Referential Retrieval and Integration in Language Comprehension: An Electrophysiological Perspective

RUNNING HEAD: Referential Retrieval and Integration

Noortje J. Venhuizen[1*], Harm Brouwer[1*]

[1]Department of Cognitive Science and Artificial Intelligence, Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands

[*]Correspondence should be addressed to: Noortje Venhuizen (n.j.venhuizen@tilburguniversity.edu) or Harm Brouwer (h.brouwer@tilburguniversity.edu)

Keywords: Event-Related Potentials; language comprehension; N400; P600; Nref

Manuscript accepted for publication in Psychological Review

Abstract

Referential processing is part and parcel of language comprehension, but in neurocognitive theories and models of comprehension it typically does not take center stage. Models informed by Event-Related Potentials (ERPs) focus on semantic and syntactic processing in terms of the two most salient ERP components, the N400 and P600, while experimental findings have implicated the Nref component—a frontal, sustained negativity—in referential processing. Extant accounts of the Nref assume it reflects processes involved in establishing reference or association at a distance, but an important open question remains how these mechanisms can be reconciled with existing neurocognitive models.

We here offer a mechanistic account of referential processing grounded in Retrieval-Integration (RI) theory, an integrated theory of language comprehension with broad empirical coverage. On RI theory, the conceptual knowledge associated with an incoming word in context is retrieved from long-term memory (N400), and accordingly integrated into the unfolding utterance representation (P600). We here argue that word meaning is not only defined by the conceptual knowledge associated with a word, but also by its referential knowledge (its presuppositions). Whenever this referential knowledge is inconsistent with what is anticipated given the context, increased referential retrieval effort ensues (Nref). In contrast to extant accounts, we do thus not implicate the Nref in the establishment of reference itself, but instead attribute referential resolution to the integrative processes underlying the P600. The resultant referential RI theory integrates the N400, Nref, and P600 in a single model, and its predictions are consistent with extant empirical evidence on referential processing.

> Mrs. Dalloway said she would buy the flowers herself.
>
> Virginia Woolf, *Mrs. Dalloway*

1 Introduction

Language comprehension is incremental in that meaning is attributed to linguistic input as it unfolds on a more or less word-by-word basis. For each word, this entails retrieving its meaning from long-term memory, and integrating it with the unfolding utterance representation that spans the entire discourse. Words may differ, however, in how they contribute to the unfolding discourse, and the contribution of the same word can be different depending on the context in which it occurs. To illustrate this, consider the following example (adapted from Boudewyn et al., 2015):

(1) The lumberjack cut down the oak and he then brought it home.

Without any context, this sentence introduces two new entities, or referents, to the discourse, a "lumberjack" and an "oak", and uses the pronouns "he" and "it" to refer back to these same entities, respectively. Now compare this to the following discourse, in which the same sentence is embedded in a larger context:

(2) *A lumberjack hiked into a forest carrying a chainsaw. He was going to cut down a tree. In a clearing he found an oak that had a mushroom on it, and an elm that had birds in its branches.* The lumberjack cut down the oak and he then brought it home.

In contrast to (1), the definite noun phrases "the lumberjack" and "the oak" in the final sentence do now not introduce novel entities, but rather refer back to entities that were previously introduced in the discourse context, rendering their contribution similar to that of the pronouns "he" and "it". While referential processing is part and parcel of online language comprehension, extant neurocognitive theories and models of comprehension, informed by Event-Related Potentials (ERPs) do not explicate the mechanisms of context-dependent referential processing.

ERPs—scalp-recorded post-synaptic neural activity time-locked to an event—offer a finegrained, multi-dimensional window into the nature and time course of cognitive processing. Of interest in the ERP signal are systematic voltage fluctuations, called components, which reflect the neural activity underlying specific operations carried out in a given neuroanatomic module or network (Näätänen and Picton, 1987). There is a spectrum of ERP components sensitive to aspects of language comprehension, including the N400, P600, Nref, and (Early) Left Anterior Negativity (ELAN/LAN)[1] (see Kutas et al., 2006; Kutas and Federmeier, 2011; van Berkum, 2009, for reviews). Of these components, the N400 and the P600 are by far the most studied, as evidenced by the increasing number of neurocognitive theories and models that aim to offer mechanistic accounts of the processes underlying these components (e.g., see Laszlo and Plaut, 2012; Rabovsky and McRae, 2014; Brouwer et al., 2017; Cheyette and Plaut, 2017; Rabovsky et al., 2018; Fitz and Chang, 2019; Li and Ettinger, 2023; Eddine et al., 2024, for computationally instantiated models). Critically, however, none of these models offer mechanistic accounts of the Nref, the component that is typically modulated in response to referential processing difficulty (van Berkum et al., 2007; van Berkum, 2009). This highlights an important shortcoming: While referential processing is critical for comprehension, it is under-represented in theories and models of language processing.

We here seek to bridge this gap by incorporating referential processing into Retrieval-Integration (RI) theory (Brouwer et al., 2012; Brouwer and Hoeks, 2013), an integrated theory of the N400 and the P600 in language comprehension with broad empirical cov-

[1]Note that the reality/robustness of the Early Left Anterior Negativity (ELAN), which may be elicited by word category violations, has been challenged, and is suggested to in fact be a baseline correction artefact (Steinhauer and Drury, 2011). Similarly, it has been suggested that the Left Anterior Negativity (LAN), which may be elicited by morphosyntactic violations, could be an artefact that arises in the grand-average from individual differences in N400 and P600 modulation patterns (Tanner, 2015).

erage (e.g., see Hoeks and Brouwer, 2014; Crocker and Brouwer, 2023) and an explicit neurocomputational instantiation (Brouwer et al., 2017, 2021b). Mechanistically, RI theory assumes that the processing of an incoming word can be conceptualized as a *process* function:

$$\textit{process: } (\textit{word form, utterance context}) \rightarrow \textit{utterance representation} \tag{1}$$

This function maps the acoustically or orthographically perceived *word form* of a word, and the *utterance context* as established after processing the preceding discourse, onto an *utterance representation*—a mental model (Johnson-Laird, 1983) or situation model (van Dijk and Kintsch, 1983; Zwaan and Radvansky, 1998)—that spans the entire discourse including the meaning contributed by the current word. Crucially, this *process* function is composed of a *retrieve* and an *integrate* function. The *retrieve* function first maps the perceived *word form* and the *utterance context* onto a representation of *word meaning*:

$$\textit{retrieve: } (\textit{word form, utterance context}) \rightarrow \textit{word meaning} \tag{2}$$

This contextualized retrieval of the conceptual knowledge associated with an incoming word is hypothesized to underlie the N400 component (see also Kutas and Federmeier, 2000; Lau et al., 2008; van Berkum, 2009)—a negative deflection in the ERP waveform that starts 200-300ms post-word onset and peaks at approximately 400ms: If the conceptual knowledge associated with an incoming word is contextually or lexically primed, its retrieval is facilitated, and N400 amplitude reduced. The *integrate* function, then, maps this retrieved *word meaning* and the *utterance context* onto an *utterance representation*:

$$\textit{integrate: } (\textit{word meaning, utterance context}) \rightarrow \textit{utterance representation} \tag{3}$$

This integration of the retrieved word meaning with the utterance context is assumed

  doi: 10.1037/rev0000530

to underlie the P600 component—a positive deflection that becomes apparent at about 600ms post-word onset: The more effort it takes to arrive at a coherent utterance representation, the higher the amplitude of the P600 (see Aurnhammer et al., 2023b, for evidence for the continuous nature of the P600). Critically, while the retrieval processes underlying the N400 are assumed to be relatively automatic, the integration processes underlying the P600 may be modulated by task demands (Kolk et al., 2003) and attention (Schacht et al., 2014).

Hoeks and Brouwer (2014) provide a first attempt to integrate the Nref findings on referential processing with RI theory. Based on the relative time-courses of the Nref and the N400, they suggest that the referential processes underlying the Nref and the lexical retrieval processes underlying the N400 run largely in parallel. Incoming words modulate both processes, such that N400 amplitude reflects the retrieval of their associated conceptual knowledge, and Nref amplitude indexes search for potential antecedents. This predicts that words may simultaneously modulate both components, but that words may differ in the degree to which they tax referential and retrieval processes: nouns, for instance, carry both referential and conceptual meaning, whereas pronouns are predominantly referential, having little conceptual meaning. While the proposal by Hoeks and Brouwer (2014) does provide a fruitful framework for thinking about referential processing within the framework of RI theory, it leaves the precise mechanisms of "searching for antecedents" rather underspecified.

We here aim to offer a mechanistic account of the processes underlying the Nref that is on par with the mechanistic accounts of retrieval (N400) and integration (P600) (Brouwer et al., 2012; Brouwer and Hoeks, 2013; Brouwer et al., 2017; Delogu et al., 2019, 2021; Aurnhammer et al., 2021, 2023a,b). Rather than assuming two parallel processes for referential search (Nref) on the one hand and lexical retrieval (N400) on the other, we put forward the hypothesis that a single retrieval process underlies these components; that is, we assume that referential knowledge, along with conceptual knowledge, directly affects the contextualized retrieval of word meaning. Critically, this retrieved word meaning may subsequently affect integration with the unfolding utterance representation.

In what follows, we will derive a referential Retrieval-Integration theory, on which the retrieval of word meaning — and its subsequent integration into the utterance representation — involves both conceptual knowledge and referential knowledge. Critically, the degree to which the unfolding discourse context primes upcoming conceptual and referential knowledge will affect the difficulty of retrieving this word meaning from long-term memory. The difficulty incurred by the establishment and/or revision of reference, then, is reflected in the general effort involved in integrating this retrieved knowledge with the unfolding utterance representation. We will derive a number of explicit predictions from referential RI theory, and describe how existing empirical findings on the electrophysiology of referential processing provide support for these predictions.

2 Referential Retrieval-Integration theory

When encountering a referential expression as linguistic input unfolds, the comprehension system is faced with resolving reference within the current linguistic and situational context. In the canonical case, referential resolution entails binding a referential expression such as a pronoun to its antecedent. For instance, in example (1), "The lumberjack cut down the oak and he then brought it home", the pronouns "he" and "it" unambiguously refer back to the discourse entities a "lumberjack" and an "oak", and hence are bound to these antecedents, respectively. Contrasting such canonical cases with cases in which referential resolution is problematic allows us to investigate how the brain responds when faced with referential difficulty, and hence to identify the neural correlates of referential processing.

To explicate the principles of referential processing, we here harness formal semantic

  doi: 10.1037/rev0000530

theory to formalize the concept of a mental model (Johnson-Laird, 1983) or situation model (van Dijk and Kintsch, 1983; Zwaan and Radvansky, 1998). Specifically, we will use the representational structures from Discourse Representation Theory (DRT; Kamp, 1981; Kamp and Reyle, 1993; Kamp et al., 2011; Venhuizen et al., 2018), a formal and representational theory of meaning developed in particular to account for anaphoric dependencies in discourse. DRT formalizes the concept of a mental/situation model in terms of recursive Discourse Representation Structure (DRS) representations. A DRS is set-theoretically defined as a set of discourse referents and a set of conditions on these referents, which may be either basic properties or relations, or complex conditions with embedded DRS contexts. DRSs can be intuitively visualized as box representations, which are meaning-equivalent to their set-theoretic representations. Example (3), for instance, gives the box-representation of the DRS for the first part of the sentence in (1).[2]

(3) The lumberjack cut down the oak

x y
lumberjack(x)
oak(y)
cut_down(x,y)

In this DRS representation, each referential expression introduces a discourse referent into the *universe* of the discourse model (x, y), and all content words introduce *conditions* on these referents into the discourse model (e.g., "lumberjack(x)"). DRSs provide the meaning of the discourse thus far, and at the same time constitute the discourse context for any upcoming information. As a discourse unfolds, its DRS representation is thus continuously updated with the contribution made by novel linguistic input (see Kamp and Reyle, 1993; Muskens, 1996; Bos, 2003). In DRT, the addition of novel information to a DRS can be formalized as a *merge* operation between two DRSs, which

[2]We here use simplified DRS representations that do not include, e.g., notions of tense and aspect. See Kamp (1981) and Kamp et al. (2011) for a comprehensive overview of the representational power of DRT. Where required, we will informally extend standard DRT syntax with propositional variables and plural entities, based on conventions in semantic theory.

is defined in such a way that the universes and conditions of both DRSs are combined (while preventing undesired variable bindings, see Kamp and Reyle, 1993 for details; see also Venhuizen et al., 2018 for a definition of merge operations for presupposed and other projected content). This operation is illustrated in (4), where the DRS from (3) is expanded with subsequent information from the sentence in (1):

(4) The lumberjack cut down the oak [and] he then brought it home

x y
lumberjack(x)
oak(y)
cut_down(x,y)

+

z v
z = x
v = y
bring_home(z,v)

=

x y z v
lumberjack(x)
tree(y)
cut_down(x,y)
z = x
v = y
bring_home(z,v)

The merge operation (denoted with '+') adds the novel information from the sentence continuation to the sentence-initial discourse context, resulting in a DRS that captures the meaning of the entire sentence. In this example, the co-referential nature of the pronouns "he" and "it" is explicitly denoted by introducing novel discourse entities (z, v) and an equality statement between the co-referring discourse referents (z = x, v = y). Note that this explicit coreference is meaning-equivalent to a DRS representation in which "he then brought it home" is formalized as "bring_home(x,y)".

DRT offers a means of directly modeling referential availability and expectedness, which we will harness to theoretically formalize a *referential* Retrieval-Integration theory. It is, however, not our aim to use DRT to generate quantitative predictions about ERP modulations due to (referential) retrieval and integration. We refer to explicit neurocomputational instantiations of RI theory (cf. Brouwer et al., 2017, 2021b) for quantitative predictions, and in the discussion we will speculate on how these can be extended to referential processing. Despite not establishing direct, quantitative linking hypotheses, it should be noted that the process of continuously merging novel information into

  doi: 10.1037/rev0000530

the discourse context is indeed conceptually similar to the mental/situation model updating assumed in RI theory, and the amount of change that needs to be made to an unfolding DRS can be thought of as the *integration* effort reflected in P600 amplitude. As for N400 amplitude, on the other hand, DRT takes the conceptual meaning associated with conditions like "cut_down(x,y)" and "bring_home(x,y)" as given, and hence does not make any predictions regarding their respective *retrieval* difficulty in context.[3] Our use of DRT is thus purely to explicate the expectedness of referential expressions, the availability of referential antecedents, and referential binding.

2.1 The referential ambiguity effect

The study by van Berkum et al. (1999) is generally considered the first investigation of the neural correlates of referential processing. They employed a design that tests the effects of *referential ambiguity* on the ERP signal. This study, which was conducted in Dutch, revealed a sustained, frontally pronounced negative deflection of the ERP signal, starting around 300ms after target onset, in response to referential noun phrases that were ambiguous in the local discourse context relative to a control condition in which the referential noun phrase unambiguously co-refers with a single antecedent from the discourse context. An example sentence for each condition is shown in (5) (translated from Dutch); the critical noun "[the] girl" (underlined below) is either ambiguous between two entities that were introduced previously ("the two girls"), as in (5-a), or refers to a single antecedent that was introduced previously in the discourse ("[the boy and] the girl"), as in (5-b). The DRT representations in (5) illustrate this context-manipulation design (here, the discourse referent 'Y' represents a plural entity for "the girls" and the DRS condition 'v=?' informally indicates referential ambiguity).

(5) a. David told the two girls to clean up their room. [. . .] David told <u>the girl</u> . . .

[3] In principle, utterance-driven retrieval difficulty could be modeled as a function of the expectancy and association of a new condition relative to the unfolding context (cf. Brouwer et al., 2017, 2021b).

x Y
x = David′
girls(Y)
tell(x, Y,“clean room”)

\+

v
x = David′
tell(x, v, ...)
girl(v)
v=?

b. David told the boy and the girl to clean up their room. [...] David told the girl ...

x y z
x = David′
boy(y)
girl(z)
tell(x, y,“clean room”)
tell(x, z,“clean room”)

\+

v
x = David′
tell(x, v, ...)
girl(v)
v = z

The frontal, sustained negativity that was observed in the ERP signal at the target word “girl” in response to the ambiguous condition (5-a) relative to the control condition (5-b), was described as a ‘Referential Ambiguity Effect’, and later became known as the ‘Nref’ effect (Nieuwland and van Berkum, 2008b). This finding was taken to indicate that “when processing the target sentence, readers very rapidly relate the noun phrase to potential referents in their representation of the earlier discourse and [...] immediately use the resulting information to parse a subsequent local structural ambiguity” (van Berkum et al., 1999, pp. 167-168). Hence, van Berkum et al. (1999) take the Nref to reflect the establishment/resolution of reference.

Critically, the referential ambiguity that occurs at the target word in (5-a) can still be resolved in the subsequent linguistic context; for example, the relative clause continuation “[David told the girl] that had been phoning to hang up” uniquely resolves the definite noun phrase to a single antecedent that is available in the discourse, with the auxiliary “had” signaling such disambiguation. As van Berkum (2009) points out, such temporary referential ambiguities are widespread and not necessarily problematic for incremental comprehension. In fact, van Berkum et al. (1999) report a small but consistent increase in P600 amplitude at the disambiguation cue "had" for 1-referent contexts (5-b) compared to 2-referent contexts (5-a), suggesting that this cue contributes more information in the 1-referent than in the 2-referent condition. In other words, when faced with a referential ambiguity, the comprehension system thus immediately anticipates further, disambiguating information. This suggests that the Nref effect is not just a marker of referential anomaly, causing the system to give up on resolution, but rather that the Nref "is reflecting the brain's natural inclination to immediately relate every shred of new information to what is known already" (van Berkum, 2009, pg. 287). Consistent with this idea are findings that report that the processes underlying the Nref are not only modulated by referential ambiguity, but also by discourse prominence (e.g., Coopmans and Nieuwland, 2020), representational richness (e.g., Karimi et al., 2018), and task demands (e.g., Nieuwland, 2014).

While the Nref *effect* is generally taken to index increased difficulty in establishing reference, there is no general consensus about the precise mechanisms underlying the Nref as a *component* (see also Hoeks and Brouwer, 2014, for discussion). van Berkum et al. note that the "fact that the Nref effect reflects something about establishing reference with respect to the situation model provides a clear constraint on its functional interpretation" (van Berkum et al., 2007, pg. 5). Furthermore, Barkley et al. link the Nref to the left anterior negativity (LAN), a phasic negativity in the 300-450ms time-window that is typically found in response to agreement violations accompanied by an increase in P600 (Molinaro et al., 2011), but that has also been implicated in referential processing (Coulson et al., 1998; Barkley et al., 2015). In particular, they suggest that "the Nref is just one instance of a family of anterior negativities elicited by association at a distance in language contexts more generally" (Barkley et al., 2015, pg. 153). None of these proposals, however, offers an explicit, mechanistic account of the precise computational operations underlying the Nref or LAN and their input and output representations; that is, none of these accounts makes it explicit if the Nref and LAN, like the N400 and P600, are modulated by every word in a sentence, nor if and how their underlying referential or "back association" processes are only selectively triggered for referential expressions. What is also unclear, is how these processes run in parallel to and potentially interact with the retrieval of word meaning underlying the N400, and what the consequences of these processes are for the integrative processes underlying the P600. To arrive at a more explicit and mechanistic explanation, we here formulate referential processing in terms of referential retrieval and integration.

2.2 Referential Retrieval

On RI theory, N400 amplitude reflects the effort involved in the contextualized retrieval of the conceptual knowledge associated with an incoming word—i.e., *word meaning*—from long-term semantic memory (see Equation 2). Grounded in theories of word meaning and concepts, the representational currency underlying this conceptual knowledge is assumed to be semantic features (for discussion, see McRae et al., 2005; Frisby et al., 2023). Retrieval is facilitated, and N400 amplitude attenuated, if through lexical or contextual priming, relevant conceptual knowledge is already (partially) activated in long-term memory. If, on the other hand, the conceptual knowledge associated with an incoming word mismatches the pre-activated knowledge, N400 amplitude increases. Conceptually, N400 amplitude can thus be seen as a measure of how consistent the meaning of an incoming word is with the current state of the long-term memory system; that is, in canonical semantic incongruities such as "He spread his warm bread with socks/butter" (Kutas and Hillyard, 1980), the larger N400 amplitude for "socks" relative to "butter" reflects the fact that the prior context "He spread his warm bread with" puts the memory system in a state that is more consistent with the conceptual knowledge associated with "butter" than with that associated with "socks". Conversely, in contrasts where the target word is equally primed across conditions such as "John [left/entered] the restaurant. Before long, he opened the menu [...]" (Delogu et al., 2019), the absence of any N400 difference reflects the fact that both "John left the restaurant" and "John entered the restaurant" put the memory system in a state that is consistent with the conceptual knowledge associated with "menu". Critically, explicit neurocomputational instantiations of RI theory (Brouwer et al., 2017, 2021b) provide support for this retrieval mechanism and its resultant explanations.

We here extend the retrieval hypothesis to the retrieval of referential knowledge. The core premise of this proposal is that the meaning of a word in context is not only defined by its associated conceptual knowledge, but also by its associated referential knowledge (its presuppositions) stored in long-term semantic memory:

$$\textit{word meaning} = \textit{conceptual knowledge} + \textit{referential knowledge} \tag{4}$$

The *retrieval* process (see Equation 2), then, can be modulated by pre-activation of both the conceptual and the referential knowledge associated with an expression. While we can only speculate about the neural representation of referential knowledge, we can start from the function that it has in referential processing, namely, to identify the presupposed referential status of the conceptual information associated with a referential expression relative to an unfolding context. This means that referential knowledge should minimally differentiate between different ways in which the conceptual information can be linked to an unfolding context, i.e., the different resolution strategies that referential expressions allow for. Following linguistic theorizing, we start from the assumption that referential expressions carry the presuppositional information that they can either refer back to a suitable and accessible referent, an operation called *binding* in DRT, or introduce a novel entity into the unfolding situation model, an operation referred to as *accommodation*. Indeed, referential expressions can be coarsely distinguished in terms of their presuppositions and resolution strategies; depending on the context, proper names and definite noun phrases can accommodate or bind to a previously introduced referent, regular pronouns typically bind to an activated antecedent, and reflexive pronouns are strictly interpreted as binding to their direct (syntactic) antecedent. Together with the conceptual knowledge associated with the expression, referential knowledge contributes to the meaning that is activated by a (referential) expression. This combined meaning, which is stored in long-term semantic memory, may be more or less activated by the unfolding context in working memory, for instance due to referential expressions that are previously introduced. Referential retrieval is facilitated in case the referential knowledge associated with an expression is pre-activated in the unfolding context, while referential retrieval difficulty—as reflected in an increase in Nref amplitude—is predicted when the referential knowledge associated with an expression mismatches with the context or is primed to a lesser degree.

To exemplify this, consider the canonical Nref-effect elicited by the referential expression "the girl" in Example (5-a) (a context that introduces "two girls") relative to Example (5-b) (a context that introduces "a boy and a girl"). The conceptual knowledge associated with "girl" should be similarly primed in both conditions, thereby eliciting no N400 difference indicative of differential difficulty in accessing the conceptual knowledge in long-term memory. The conditions critically differ, however, in terms of the referential knowledge that is activated by the context; the introduction of "two girls" in (5-a) primes away from referential expressions that presuppose the existence of a unique female antecedent, whereas the context in (5-b) activates this referential knowledge. Hence, the referential knowledge associated with the definite, singular noun phrase "the girl"—the presupposition that there is a unique referent in the discourse context that constitutes a suitable antecedent for the expression that matches in terms of its associated conceptual knowledge (cf. Van der Sandt, 1992; Geurts, 1999)—is activated to a lesser degree in (5-a) compared to (5-b), thereby increasing referential retrieval difficulty as reflected in an increase in Nref amplitude.

  doi: 10.1037/rev0000530

In sum, the critical extension to the retrieval hypothesis is thus that word meaning includes referential knowledge, and that referents in the unfolding situation model in working memory may prime this knowledge to a larger or lesser degree in long-term memory. Nref amplitude, then, is a measure of how consistent the referential meaning of an incoming word is with the current state of the long-term memory system.

2.3 Referential Integration

On RI theory, retrieved *word meaning* needs to be integrated with the current *utterance context* to produce an updated *utterance representation* (see Equation 3). Indeed, difficulties in retrieving conceptual knowledge from long-term memory may entail semantic integration difficulty, as indexed in increased P600 amplitude, since unanticipated conceptual knowledge will typically also have utterance-level consequences; e.g., what does it mean for someone to "spread his warm bread with socks?" (Kutas and Hillyard, 1980; see Aurnhammer et al., 2023a for evidence at the single-trial level). Similarly, difficulties in retrieving referential knowledge may also entail referentially-induced integration difficulty, and hence increase P600 amplitude, reflecting problems in establishing referential coherence.

Canonical examples of increased integration difficulty due to referential resolution are cases in which novel discourse entities need to be accommodated. For instance, in the contrast "Tobias visited a [concert/conductor] in Berlin. He said that the conductor ...", a P600-effect ensues at "the conductor" when this entity is not previously introduced into the discourse (Burkhardt, 2006). Critically, we do not predict a (large) difference in Nref and/or N400 amplitude here, as the referential and conceptual knowledge associated with the definite noun phrase "the conductor" is primed in both conditions: the introduction of "concert" comes with a strong presupposition about the existence of a uniquely identifiable conductor, to a similar degree as the explicit introduction of "a conductor" carries this presupposition (see also Burkhardt, 2007; Schumacher, 2011).

  doi: 10.1037/rev0000530

In case of referential ambiguity, in turn, increased integration difficulty due to establishing reference is predicted to ensue only when it becomes apparent that either the ambiguity cannot be resolved (leading to referential failure), or that the referential expression needs to be accommodated by introducing a new entity. This explains why in Example (5-a), no integration difficulty ensues at the critical referential expression, relative to (5-b), as evidenced by the sustained Nref-effect and the absence of a clear P600 difference. As discussed above, while the referential expression "the girl" is referentially ambiguous, it does not directly induce a revision of the utterance representation constructed so far, as the ambiguity can still be resolved post-nominally; e.g., with a relative clause continuation like "the girl, who had been phoning …". Indeed, van Berkum et al. (1999) show that when the sentence is instead continued with a complement clause continuation that does not resolve the ambiguity (e.g., "[David told the girl] that there would be some visitors"), an increase in P600 amplitude at the expletive "there" is observed for ambiguous contexts compared to unambiguous contexts, indicating integration difficulty due to failure to establish reference.

2.4 Predictions

Referential Retrieval-Integration theory predicts that each incoming word modulates conceptual (N400) and referential retrieval effort (Nref), as well as integration effort (P600). We here focus on how referential retrieval and integration is modulated by explicit referential expressions (see Brouwer et al., 2012, for general predictions from RI theory regarding the N400 and the P600). The referential retrieval hypothesis predicts that retrieval difficulty is modulated by the degree to which the referential and conceptual knowledge associated with a referential expression is consistent with the prior context; that is, lexical and contextual priming (e.g., the explicit introduction of referents in the discourse context) may lead to pre-activation of the referential and conceptual knowledge associated with a referential expression, thereby attenuating Nref and

Table 1: **Predictions From Referential Retrieval-Integration theory.**

Referential Expression			Referential retrieval	Referential integration
# Antecedents	**Status**	**Resolution**		
1	unambiguous	bind	–	–
≥ 2	ambiguous	?	Nref	–
≥ 2	unresolvable	fail	–	P600
1	unexpected	bind	Nref	–
1	underspecified	bind	(Nref)	P600
0	novel	accommodate	(Nref)	P600
0	incompatible	fail	–	P600

N400 amplitude. As noted above, different types of referential expressions may vary in terms of the referential knowledge that is associated with them (e.g., reflexive pronouns convey referential knowledge that codes for binding, whereas proper names primarily convey knowledge that codes for accommodating novel referents), as well as in terms of their conceptual knowledge (e.g., definite noun phrases typically are conceptually richer than pronouns). As a result, prior activation that is due to priming from the unfolding utterance context may differentially affect the processing of incoming referential expressions. Furthermore, it should be noted that the strength of this priming effect is modulated by the temporal distance between the introduction of an antecedent and the referential expression (Chow et al., 2018; Nakamura et al., 2024). Table 1 summarizes the different constellations for the availability and status of referential antecedents, the resolution strategies for referential expressions, and the predicted component modulations for referential retrieval and integration. In what follows, we will discuss these constellations and predictions in detail.

When a referential expression matches the preactivated knowledge—e.g., the expression is *unambiguous* with respect to its antecedent—retrieval is facilitated (see Table 1). In the canonical referential ambiguity cases, the conceptual knowledge associated with an ambiguous referential expression (e.g., "the girl") is primed by the prior introduction of multiple potential antecedents ("two girls"), but referentially these multiple antecedents prime away from referential knowledge that identifies a uniquely iden-

  doi: 10.1037/rev0000530

tifiable referent. Hence, referential retrieval effort is predicted to increase upon encountering an *ambiguous* referential expression. Critically, the degree of referential retrieval difficulty, and hence Nref amplitude, is predicted to be a function of how suitable and accessible antecedents are; that is, suitability and accessibility may be affected by factors such as recency, prominence, and salience of the antecedents, which will differentially prime referential knowledge, e.g., pertaining to activated referents. Nref amplitude is thus not a direct reflection of the amount of referential uncertainty per se. Indeed, increased referential retrieval difficulty may also ensue in unambiguous contexts, when the referential knowledge associated with a referential expression mismatches with the referential knowledge that is primed by the discourse context; that is, when the referential status of an expression is *unexpected* (e.g., a proper name is used instead of a pronoun).

A referential expression will induce integration difficulty, and hence increase P600 amplitude, whenever it leads to a revision of the unfolding situation model, e.g., due to *underspecified* discourse referents, the introduction of a *novel* discourse referent or when the expression is *incompatible* because referential cohesion cannot be achieved. In situations in which the unfolding situation model biases towards particular referential continuations (e.g. in terms of gender), an unexpected referential expression may additionally induce increased referential retrieval difficulty due to reduced priming of the referential knowledge associated with the referential expression. Conversely, when faced with an *ambiguous* referential expression, no increased integration difficulty reflecting referential inference or failure is predicted; that is, in these situations integration of the plausible but ambiguous referential expression leads the system to postpone binding or accommodation and to await further disambiguating information, thereby not increasing P600 amplitude, but rather leading to a sustained Nref. However, if the ambiguous referential expression is perceived as *unresolvable*, for instance because later disambiguation is perceived as unlikely, an increase in P600 amplitude is predicted, reflecting the integration difficulty incurred by the failure to achieve referential cohesion. Whether or not a referential expression is perceived as unresolvable depends on the type of referential expression, and may also be subject to individual differences, as well as task demands.

Finally, it should be noted that while we here focus on referential expressions, we believe the referential retrieval hypothesis can be generalized to any situation model content that incoming information can refer back to; that is, a verb assigning roles to specific constituents, for instance, may be more or less consistent with the state of long-term memory depending on what constituents are processed thus far, and how the resultant situation model in working memory primes upcoming information, similar to how a pronoun can be more or less consistent with the state of long-term memory depending on the referents encountered thus far. Indeed, this would be consistent with the proposed link between the Nref and the LAN (Barkley et al., 2015), and explain the LAN that is found in response to agreement violations in terms of retrieval.[4]

3 The neural correlates of referential processing

Following the predictions from referential RI theory outlined in Table 1 regarding the neural correlates of referential processing, we here focus on studies investigating referential processing using electroencephalography in situations in which either multiple antecedents are available (2-ref), there is a single antecedent with attenuated referential accessibility (1-ref), or no antecedents are available (0-ref). We show that the empirical

[4] If one accepts the retrieval view on the Nref and the LAN, an open question remains why these components are morphologically distinct; that is, while the Nref and the LAN are both frontal negativities, the Nref is sustained, and the LAN phasic and left-pronounced. We believe these morphological differences arise because in the case of an ambiguous pronoun, further information may help disambiguate, thereby not increasing integrative processing difficulty at the verb, whereas in the case of an agreement violation, the infelicity is clear at the verb, leading to increased integrative processing difficulty. Indeed, unlike the sustained Nref, the LAN (like the N400) is typically accompanied by an increase in P600 amplitude, rendering it phasic, and through spatiotemporal component overlap potentially more left-dominant (see Brouwer and Crocker, 2017, for a discussion of component overlap).

evidence is consistent with the predictions from referential RI theory, and discuss how differences between types of referential expressions, as well as individual differences between participants, may result in varying resolution strategies that lead to differential modulation of the ERP components.

3.1 Multiple available referents (2-ref)

The canonical Nref effect, reported by van Berkum et al. (1999), manifests as a frontal, sustained negativity in response to referential expressions in ambiguous contexts with multiple available antecedents, illustrated in (5-a), relative to contexts with a uniquely available antecedent (5-b). This effect has been replicated in different modalities (auditory: van Berkum et al., 2003; written: Nieuwland et al., 2007a), across different languages (Dutch: Nieuwland et al., 2007a; Spanish: Martin et al., 2012, 2014; English: Boudewyn et al., 2015) and for different referential expressions (definite noun phrases: Boudewyn et al., 2015; determiners: Martin et al., 2012, 2014; pronouns: Nieuwland and van Berkum, 2006). Table 2 provides an overview of various studies that directly investigate referential ambiguity.

3.1.1 Referential ambiguity

Referential RI theory predicts an increase in referential retrieval effort, reflected in the Nref component, in case the referential knowledge associated with a referential expression is less anticipated. In standard referential ambiguity manipulations, the availability of multiple antecedents is incompatible with the referential knowledge associated with a uniquely selecting referential expression. Critically, the Nref component is not only modulated by the *presence* of multiple suitable referents in the discourse context, but also by the *availability* of these referents to serve as antecedents for the current referential expression. Nieuwland et al. (2007a), for instance, tested effects of ambiguity in cases where multiple suitable referents are introduced, but the availability of these

Table 2: **2-Ref: Referential Ambiguity.** Lang. = language. Task abbreviations: Compr. = Comprehension questions (typically for a percentage of trials); None = No task (read for comprehension). Language abbreviations: DUT = Dutch; ENG = English; SPA = Spanish. Nref/N400/P600 modulations: "+" = significant effect reported; "−" = no significant effect reported; "(+)" = numeric difference. Studies marked with asterisk: auditory modality.

Study	Manipulation	Task	Lang.	Example: [Target/Control]	Nref	N400	P600
vBBH (1999), vBBHZ (2003)*	Ambiguity	None	DUT	David had told [the two girls/the boy and the girl] to clean up their room. But [**one of the girls**/the boy] had stayed in bed all morning, and [**the other**/**the girl**] had been on the phone all the time. David told the girl …	+	−	−
N&vB (2006)	Ambiguity	None	DUT	**Al Pacino** told [**Bruce Willis**/ Madonna] that he …	+	−	−
NOvB (2007a)*	Ambiguity	None	DUT	[…] Jim had been talking to **his nephew** who was very much into politics and [**another one**/his uncle] who was really into history… The nephew …	+	−	−
	Availability	None	DUT	…**The nephew** who was into history [kept telling stories/ left early], and/but **the nephew** who was into politics (also) kept rambling on. He told the nephew …	+	−	−
N&vB (2008a)	Ambiguity	None	DUT	Britney Spears had several pieces of jewelry, including **a golden necklace** and **a silver [one/bracelet]** … She was admiring the necklace …	+	−	(+)[a]
MNC (2012,2014)	Embedding	Compr.	SPA	Marta bought the **t-shirt**$_F$ that was next to the **skirt**$_M$ and Miren took [another$_M$/another$_F$] …	+	−	−
BLTLDMCS (2015)	Ambiguity	Compr.	ENG	A lumberjack […] found **an oak** that had a mushroom on it, and [**an oak**/an elm] that had birds in its branches. The lumberjack cut down the oak …	+	−	−
FCG (2018)	Ambiguity	Compr.	ENG	**Tyler**$_M$/Janet$_F$ astonished **Eric** because he …	(+)[b]	−	(+)[b]
KSF (2018)	Ambiguity	Compr.	ENG	**The actor** (who was visibly upset) walked away from [**the cameraman**/the actress] (who was critical of the show). After a while, he …	(+)[c]	−	−

[a]Splitting participants based on the sign of the mean difference between the conditions (posterior channels, 500-1600ms tw) reveals two groups of participants: positivity-to-ambiguity group (n=19): P600; negativity-to-ambiguity group (n=20): Nref.

[b]No significant main effects, participants grouped following Nieuwland and van Berkum (2008a): negativity-to-ambiguity group (n=18): Nref—related to Working Memory/Count Span; positivity-to-ambiguity group (n=17): P600—marginally negatively related to Attention/Stroop.

[c]Trend toward significance for main effect of Ambiguity; No significant interaction between Ambiguity and Richness (Karimi et al., 2018, p. 76).

 doi: 10.1037/rev0000530

referents is manipulated, as illustrated in (6):

(6) *Context:* At the family get-together, Jim had been talking to one nephew who was very much into politics and another one who was really into history. [...]

a. The nephew who was into history kept telling boring stories, and the other one also kept rambling on. [...] He$_{\text{[Jim]}}$ told the nephew ...

x y[AVAILABLE] z[AVAILABLE]
x = Jim′
nephew(y)
nephew(z) ¬ [y=z]
loves(y,'politics')
loves(z,'history')
tell_stories(z)
ramble_on(y)

\+

v[AVAILABLE]
x = Jim′
tell(x, v, ...)
nephew(v)
v = ?

b. The nephew who was into history left early, but the nephew who was into politics kept rambling on. [...] He$_{\text{[Jim]}}$ told the nephew ...

x y[AVAILABLE] z
x = Jim′
nephew(y)
nephew(z) ¬ [y=z]
loves(y,'politics')
loves(z,'history')
leave(z)
ramble_on(y)

\+

v[AVAILABLE]
x = Jim′
tell(x, v, ...)
nephew(v)
v = y

In the ambiguous condition (6-a), the target word "[the] nephew" has two suitable and accessible antecedents, namely the nephew that is into politics and the one that is into history, which are equally available/in focus. This is indicated in the DRT representation of (6) by supplementing the referential information (described in the *DRS universe* that is shown in the top part of the DRS) associated with both the discourse context and the referential expression with explicit features for referential availability. By contrast, in the control condition (6-b), the discourse context results in one of these antecedents becoming unavailable (though referentially still accessible), because the potential referent has 'left the scene'. In terms of referential RI theory, this yields a constellation that is similar to the standard referential ambiguity cases: the referential information associated with the singular simple noun phrase—that the referent is uniquely identifiable and available—is more anticipated in the control condition (6-b) than in the ambiguous condition (6-a). Indeed, Nieuwland et al. (2007a) report an Nref effect similar to the standard referential ambiguity cases in response to (6-a) as compared to (6-b).

3.1.2 Individual differences

In standard manipulations of referential ambiguity, the referentially ambiguous target word is typically an expression for which a disambiguating cue is anticipated to resolve the (temporary) ambiguity (e.g., a relative clause as in van Berkum et al., 1999, or a noun as in Martin et al., 2012, 2014). In these constellations, an Nref is predicted to ensue because of a mismatch between the referential knowledge associated with the ambiguous target word and the context. In the case of ambiguous bare pronouns, however, it is less clear whether the subsequent context can offer such a disambiguating cue. This potentially results in increased integration difficulty reflecting revision of the current situation model to resolve or attune to the elicited ambiguity (e.g., by introducing a novel referent), which according to RI theory is reflected in an increase in P600 amplitude (Brouwer et al., 2012). Indeed, the literature has reported evidence for both Nref and P600 responses to referential ambiguities, both within as well as across studies (Nieuwland and van Berkum, 2006, 2008a; Fiorentino et al., 2018, see Table 2 for an overview).

Fiorentino et al. (2018), for instance, conducted a simple 2-referent vs. 1-referent study, e.g., "Tyler$_{M}$/ Janet$_{F}$ astonished Eric because he …", but found no statistical dif-

  doi: 10.1037/rev0000530

ferences between conditions in the 500-1500ms (Nref) time window (visual inspection also reveals no differences in the standard P600 time window, 600-1000ms). However, they argue that following Nieuwland and van Berkum (2008a), their participants can be divided into two subgroups based on their response to the ambiguity: a *Negativity-to-Ambiguity* group that shows a significant sustained negativity (Nref) in response to the 2-referent condition, relative to the 1-referent condition, and a *Positivity-to-Ambiguity* group that shows a significant late positivity (P600) for the 2-referent compared to the 1-referent condition. Analyses of individual differences between the groups observed by Fiorentino et al. (2018) reveal that within the *Negativity-to-Ambiguity* group, the size of the Nref effect is related to working memory measures (Conway et al., 2005), with participants with a higher counting span score showing larger negativities. For participants in the *Positivity-to-Ambiguity* group, on the other hand, the effect was marginally related to measures of attention control (Bush and Shin, 2006), with participants that performed better on the Stroop task showing smaller positivities.

These findings support the idea that referential ambiguity may have differential effects on a by-item and by-participant basis. Since the anticipation of referential knowledge is modulated by the cohesiveness and richness of the situation model in working memory, and these situation models may differ between participants, the degree to which referential knowledge is activated/primed may consequentially also differ. Specifically, participants may construct a situation model in which two or more referents are equally salient/in focus, which primes away from referential knowledge associated with expressions that presuppose a unique, activated antecedent (e.g., pronouns). Indeed, this explains the increase in Nref amplitude for the *Negativity-to-Ambiguity* group. The relation between working memory measures and the size of the Nref effect reported by Fiorentino et al. (2018) is in line with the hypothesis that an increase in the Nref component reflects a mismatch between the referential meaning of an incoming word and the unfolding utterance representation in working memory.

Alternatively, if the situation model that is being constructed imposes a bias towards a specific referential expression, encountering an ambiguous expression will result either in a revision of the unfolding situation model (e.g., updating the status of the available referents) or a failure to achieve referential coherence, which should both be reflected in an increase in P600 amplitude. This bias may result from, for instance, an interpretational preference toward the subject due to the verb "astonished" (similar to implicit causality verbs; van Berkum et al., 2007, see section 3.2.2). Furthermore, given the dependence of the P600 on task demands (Kolk et al., 2003; Schacht et al., 2014; see Brouwer et al., 2012 for discussion), the integration difficulty resulting from this revision is also predicted to be modulated by the task at hand, which may explain the marginally significant effect of attention control in the *Positivity-to-Ambiguity* group. Finally, as different task demands may lead to differences in depth of processing, this is also predicted to have an effect on the referential knowledge that is primed by the situation model under construction.

3.2 Single available referent (1-ref)

While the Nref effect is traditionally associated with cases of referential ambiguity, more recent work has reported effects with the same scalp distribution and latency for non-ambiguous contrasts. Indeed, according to the referential Retrieval-Integration hypothesis, referential expressions are predicted to elicit increased referential retrieval difficulty whenever the referential knowledge associated with the expression mismatches with the referential status of the utterance context; this is not only the case when multiple antecedents are available, but also when the type of referential expression is unexpected (e.g., a pronoun is used instead of a proper name, or vice versa). Furthermore, referential expressions may induce increased integration effort in case the conceptual knowledge associated with the expression needs to be integrated into the unfolding utterance representation. Below, we describe these referential configurations in more

  doi: 10.1037/rev0000530

detail and discuss relevant experimental findings.

3.2.1 Referential expectedness

Table 3 provides an overview of electrophysiological studies that focus on manipulations in which the *referential knowledge* that is associated with a referential expression is differentially pre-activated by the utterance context; for instance, by manipulating the *prominence* of antecedents for pronouns and proper names (Swaab et al., 2004), the *distance* between a pronoun and its antecedent (Streb et al., 2004; Hammer et al., 2008), or the *richness* of the situation model described by the context (Karimi et al., 2018).

In the study by Karimi et al. (2018), both representational richness and ambiguity were manipulated, and a main effect of richness was observed. Critically, they show that for unambiguous sentences, in which the pronoun unambiguously co-refers with one of the antecedents available in the discourse context, an Nref effect is elicited for a minimal condition, shown in (7-a), relative to a representationally rich condition, shown in (7-b).

(7) a. The actor walked away from the actress. After a while, he...

x y
actor(x)
actress(y)
walk_from(x,y)

\+

z[FOCUS]
male(z)
z=x?

b. The actor who was visibly upset walked away from the actress who was critical of the show. After a while, he ...

x[FOCUS] y[FOCUS]
actor(x)
visibly_upset(x)
actress(y)
critical_of_show(y)
walk_from(x,y)

\+

z[FOCUS]
male(z)
z=x

While in the minimal condition (7-a), there is no ambiguity or referential uncertainty,

Table 3: **1-Ref: Referential Expectedness.** Lang. = language. Task abbreviations: Compr. = Comprehension questions (typically for a percentage of trials). Language abbreviations: ENG = English; GER = German. Nref/N400/P600 modulations: "+" = significant effect reported; "−" = no significant effect reported; "(+)" = numeric difference. Studies marked with asterisk: auditory modality.

Study	Manipulation	Task	Lang.	Example: [Target/Control]	Nref	N400	P600
SCG (2004)	Prominence	Compr.	ENG	**John** [∅/and Mary] went to the store so that he …	−	−	−
	Prominence	Compr.	ENG	**John** [∅/and Mary] went to the store so that John …	(+)	+	−
SHR (2004)	Distance	Compr.	GER	[s1/s3:] **Anna** wants to go on a walking-tour. s2: Gerhart is an experienced mountaineer. [s3/s1]: The weather is beautiful today. s4: Then she …	(+)	+	−
	Distance	Compr.	GER	s1: The weather is beautiful today. [s2/s3:] **Anna** wants to go on a walking-tour. [s3/s2]: Gerhart is an experienced mountaineer. s4: Then she …	(+)	+	−
LGCS (2007), CLBGS (2007)*	Prominence	Compr.	ENG	At the office, **Daniel** [∅/and Amanda] moved the cabinet because Daniel …	(+)	+	−
	Repetition	Compr.	ENG	At the office, **Daniel** moved the cabinet because [Daniel/ Robert] …	−	−	−
	Repetition	Compr.	ENG	At the office, **Daniel** and Amanda moved the cabinet because [Robert/Daniel] …	(+)	+	−
HJLM (2008)	Distance	Compr.	GER	The **chief** [attacks soon and/∅] is martial, because he …	(+)	+	−
	Distance	Compr.	GER	The **apple**$_{M}$ [is very juicy and/∅] is sweet, because he …	−	−	−
KSF (2018)	Richness	Compr.	ENG	**The actor** [∅/who was visibly upset] walked away from the actress (cameraman) [∅/who was critical of the show]. After a while, he …	+	−	−

the observed Nref effect in response to the minimal condition, compared to the representationally rich condition (7-b), suggests that in the latter condition, the referential knowledge associated with the pronoun "he" is primed to a stronger degree. We hypothesize that this is because in the representationally rich condition the antecedent of the pronoun is put 'in focus', while this is not the case in the minimal condition, as illustrated in the DRS representations above. These results show referential retrieval is not just modulated by referential ambiguity, but rather by more general referential expectedness. Indeed, while Karimi et al. (2018) also report an increased negativity for ambiguous versus unambiguous pronouns in both of these contexts (trending towards significance), this effect appeared to be smaller than the effect of representational richness, thus suggesting that ambiguity may simply be another factor affecting referential expectedness and thereby referential retrieval difficulty.

The finding that representational richness facilitates referential retrieval may appear to be somewhat at odds with findings from studies that manipulate the *distance* between a referential expression and its antecedent; referential expressions whose antecedents are introduced further away in the discourse typically induce increased processing difficulty, as compared to referential expressions whose antecedents are close (see Table 3). Hammer et al. (2008), for instance, found that congruent pronouns that refer back to a person (but not those that refer back to an object/thing) elicit a broadly distributed negativity in the 200-700ms time window when there is more intervening information between the antecedent and the pronoun, compared to when there is less information between the antecedent and the pronoun. Similarly, Streb et al. (2004) show increased sustained negativities for pronouns that co-refer with antecedents introduced in the first (far) and second (middle) sentence, compared to the third (near) sentence of a four-sentence discourse. While these effects appear to have different scalp distributions than the traditional frontally-pronounced Nref effects in response to referential ambiguities, these findings do suggest increased referential retrieval difficulty when the primed referential knowledge of the antecedents decays over time due to intervening material (see Chow et al., 2018; Nakamura et al., 2024, for similar recency effects on the priming of conceptual knowledge), suggesting a trade-off for referent retrieval with respect to representational richness (i.e., focus/salience of antecedents) and antecedent recency.

Studies that investigate the effect of referential *repetition*—specifically with respect to proper names—can be seen as instances of referential expectedness manipulations; repeating a proper name in contexts in which there is only a single referential entity available (e.g., "John went to the store. John ...") has been shown to result in an increased negativity compared to (unambiguous) conditions in which multiple named referential entities are available (e.g., "John and Mary went to the store. John ..."; Swaab et al., 2004; Ledoux et al., 2007; Camblin et al., 2007). This effect, which has become known as the Repeated Name Penalty, is typically interpreted as an N400 effect (see, e.g., Swaab et al., 2004). However, this conclusion is difficult to reconcile with the retrieval hypothesis, as the conceptual information associated with the proper name should already be activated in both contexts. Instead, it has been suggested by Hoeks and Brouwer (2014) that the Repeated Name Penalty can also be interpreted as an Nref effect. Indeed, on the referential retrieval hypothesis, an Nref effect is predicted in cases where the referential information associated with a referential expression is less anticipated. In contexts that introduce a single, suitable and activated antecedent, the conceptual information associated with this antecedent is primed by the situation model, but referentially, this situation model primes away from the same proper name; rather, it primes information consistent with the use of a pronoun for this antecedent (in contrast to contexts in which multiple named discourse referents are available). Hence, the Repeated Name Penalty is taken to be an instance of the Nref effect which reflects increased referential retrieval difficulty due to the context priming away from the repeated use of proper names to refer to the same (activated) entity.

In summary, according to the referential instantiation of RI theory, when the use of a

  doi: 10.1037/rev0000530

referential expression in unambiguous contexts is unexpected due to attenuated accessibility or mismatching referential information (as in the case of repeated names), this leads to increased referential retrieval effort reflected in the Nref component. Accessibility can be affected by (lack of) representational richness, i.e., a richer representation of the unfolding utterance/discourse meaning may result in increased pre-activation of relevant referential information and therefore reduced referential retrieval difficulty. Conversely, antecedents may become less accessible (less primed) when intermediate semantic information is introduced in the discourse context, which may interfere with retrieval of the referential information.

3.2.2 Referential inference

Even when a referential expression unambiguously selects an antecedent from the discourse context, increased processing difficulty may nevertheless ensue in integrating the conceptual meaning associated with the referential expression in the unfolding utterance representation. Specifically, referential RI theory predicts that whenever the unfolding representation of the utterance meaning needs to be revised—or augmented, see Section 3.3 below—due to inferencing, an increase in integrative processing, reflected in enhanced P600 amplitude, ensues. Table 4 illustrates a variety of studies in which unambiguous referential expressions trigger inferential reasoning with regard to the unfolding utterance representation; in all of these studies, existing interpretation biases (due to, for instance, causal structure, gender stereotypes or syntactic embedding) are disconfirmed by the introduction of a referential expression, which results in a revision of the unfolding utterance representation.

van Berkum et al. (2007), for instance, investigated pronoun interpretation in the context of interpersonal "implicit causality" verbs, which are verbs that describe events between multiple individuals and that convey a bias towards the immediate cause of the event at hand (Garvey and Caramazza, 1974). This is illustrated in (8), where the

Table 4: **1-Ref: Referential Inference.** Lang. = language. Task abbreviations: Accep. = Acceptability judgments (typically every trial); Compr. = Comprehension questions (typically for a percentage of trials); None = No task (read for comprehension). Language abbreviations: CHI = Chinese; DUT = Dutch; ENG = English; GER = German. Nref/N400/P600 modulations: "+" = significant effect reported; "−" = no significant effect reported; "(+)" = numeric difference. Studies marked with asterisk: auditory modality.

Study	**Manipulation**	**Task**	**Lang.**	**Example: [Target/Control]**	**Nref**	**N400**	**P600**
OBM (1997)	Gender bias	Accep.	ENG	The **nurse** prepared [himself/herself] …	−	−	+
vBKON (2007)	Causal. bias	None	DUT	[**David** praised Linda/Linda praised **David**] because he...	−	−	+
MBS (2007)	Concreteness	Accep.	GER	Although the **Titanic** was said to be unsinkable, it went down. This [accident/ship] …	−	−	+
L&Z (2010)	Verb bias	Compr.	CHI	**Xiaoli** asked **Xiaozhang** not to [embroil/disguise] ziji …	−	−	+
W&Y (2013)	Congruency	Accep.	CHI	Laowei and Laowu are both colleague teachers. **Laowei** teaches music, whereas **Laowu** teaches painting. A student brought his [composed song/painting] to Laowu …	−	+	+
CGO (2015)	Gender bias	Accep.	ENG	The **architect** saw [herself/himself] …	+	−	+
MGSGF (2015)*	Embedding	Compr.	GER	During the rehearsal, **the conductor**$_{M}$, that **the pianist**$_{F}$ irritated, interrupted the prelude and [she/he] …	−	−	(+)[a]
Xu (2015)	Embedding	Compr.	CHI	**Xiaolan**$_{F}$, **Xiaoming**$_{M}$ know, [he/she] …	−	−	+

[a]Numeric positivity based on visual inspection of figures in the Supplementary Materials of Meyer et al. (2015); no significant effect in traditional ERP analysis reported over complete time window (0-1s).

verb "praise" biases towards an explanation that involves the object of the *praise*-event, i.e., the individual referred to in the second argument of the verb (represented as the focused referential entity y in the DRS representations below):

(8) a. David praised Linda because <u>he</u>...

b. Linda praised David because <u>he</u>...

(a)

x y[FOCUS] p
David(x) Linda(y) praise(x,y) cause(praise,p) p: [v[FOCUS] | male(v), ~~v=y~~ **v=x**, ...]

(b)

x y[FOCUS] p
Linda(x) David(y) praise(x,y) cause(praise,p) p: [v[FOCUS] | male(v), **v=y**, ...]

In the target condition shown in (8-a), the verb "praise" leads to a situation model in which "Linda" is anticipated to be the direct cause of the *praise*-event. Encountering the pronoun "he", which unambiguously refers to the individual introduced as the first argument of the verb "praise" (namely, David), thus disconfirms this anticipation and therefore results in a revision of the utterance representation constructed so far (illustrated in the DRS above by striking through the anticipated reference assignment in the complement clause of "because"). In the control condition shown in (8-b), on the other hand, no revision of the implicit causality captured in the utterance representation is predicted, since the pronoun "he" matches the implicit causal inference of the verb "praise". Indeed, van Berkum et al. (2007) report a P600 effect in response to bias-inconsistent pronouns as in (8-a) when compared to bias-consistent pronouns as in (8-b).

Similar to how implicit causality may lead to interpretation biases, gender stereo-

  doi: 10.1037/rev0000530

types may also result in a biased interpretation that can be either confirmed or disconfirmed by the use of a gendered referential expression. For instance, Osterhout et al. (1997) compared target sentences like "The nurse prepared himself [...]" (where "nurse" has a female bias) to control sentences like "The nurse prepared herself [...]" and observed an increase in P600 amplitude at the reflexive pronoun "himself" (though the effect was smaller than in definitional gender violations, which are discussed in Section 3.3 below). This effect was replicated by Canal et al. (2015), who in addition to the P600, which was more pronounced in parietal electrodes, report a frontally distributed Nref effect. Indeed, referential RI theory predicts that a revision of the antecedent that is due to an initial gender bias should be reflected in an increase in P600 amplitude. The observed Nref effect suggests that the combined conceptual and referential information that is associated with the gender-biased context differentially primes for different reflexive pronouns. That is, the interpretation bias that is introduced by a gender-stereotyped word like "nurse" may have a direct effect on the referential information that is primed in this context, activating female pronouns to a stronger degree than male pronouns. This is consistent with Nref effects observed in studies that manipulate definitional gender mismatch (e.g. Nieuwland, 2014, see Section 3.3 below). However, since we expect the activation of referential knowledge to be influenced by many factors, including individual differences and task constraints, we can only speculate on why the frontal negativity was not observed in earlier studies such as Osterhout et al. (1997); we refer to Canal et al. (2015) and Nieuwland (2014) for a more elaborate discussion of the inconsistency in this particular set of results.

In sum, a referential expression that unambiguously selects for an existing antecedent may cause a revision of the situation model constructed so far due to a mismatch between the referential expression and interpretation biases in the situation model. These interpretation biases may be the result of lexical-level biases (implicit causality, gender), as well as grammatical (subject-hood, embedding) or discourse-level biases (coherence;

cf. Kehler et al., 2008). The mismatch between this interpretation bias and an incoming referential expression triggers a revision of the utterance representation constructed so far, which results in integrative processing difficulty indexed by an increase in P600 amplitude. In addition, interpretation biases may lead to differential activation of referential information (depending on, for instance, task constraints), which in turn may lead to differences in referential retrieval difficulty, as indexed by Nref amplitude.

3.3 Unavailable referents (0-ref)

Referential expressions are not only used in contexts in which they refer back to previously introduced antecedents, but may also be used in contexts where no suitable antecedent is available. Depending on the context and the type of referential expression, this either results in the introduction of novel referents, or in a failure to establish referential cohesion. Since both of these constellations result in a revision or augmentation of the utterance representation constructed so far, referential RI theory predicts increased integration difficulty in both cases, as reflected in P600 amplitude.

3.3.1 Referent introduction

Early studies on referential processing during language comprehension focus on the *availability* of referential antecedents, typically by manipulating the definitional gender (mis)match between a referential antecedent and a pronoun (see Table 5). Osterhout and Mobley (1995), for instance, presented participants with the following sentences (in isolation, so without any linguistic context):

Table 5: **0-Ref: Referent Introduction.** Lang. = language. Task abbreviations: Accep. = Acceptability judgments (typically every trial); Compr. = Comprehension questions (typically for a percentage of trials); Fix = Instruction to fix referential failure (before experiment); None = No task (read for comprehension); Verif. = verification questions (typically every trial). Language abbreviations: CHI = Chinese; DUT = Dutch; ENG = English; GER = German; SPA = Spanish. Nref/N400/P600 modulations: "+" = significant effect reported; "−" = no significant effect reported; "(+)" = numeric difference.

Study	Manipulation	Task	Lang.	Example: [Target/Control]	Nref	N400	P600
O&M (1995)	Availability	Accep.	ENG	**The aunt** heard that [he/she] …	−	−	+
SLM (2002)	Availability	Compr.	GER	**The boy**$_{M}$ wants to sleep and therefore [she/he]…	−	+	+
	Availability	Compr.	GER	**The little boy**$_{NEUT}$ wants to sleep and therefore [she/it/he]..	−	−	+
LJHM (2006)	Availability	Compr.	GER	**The friend**$_{M}$ was invited and therefore they did count on [her/him]…	−	+	+
	Availability	Compr.	DUT	**The friend**$_{M}$ was invited and therefore they did count on [her/him]…	−	+	+
N&vB (2006)	Availability	None	DUT	Al Pacino told [Bruce Willis/**Madonna**] that she …	−	−	+
B (2006)	Bridging	Verif.	GER	Tobias visited [a **concert**/a **conductor**] in Berlin. He said that the conductor …	−	−	+
HJLM (2008)	Thing:Avail.	Compr.	GER	**The apple**$_{M}$ is sweet, because [she/he] …	−	−	+
	Person:Avail.	Compr.	GER	**The chief**$_{M}$ is martial, because [she/he] …	−	+	−
	Thing:Avail.	Compr.	GER	**The apple**$_{M}$ is very juicy and is sweet, because [she/he]...	−	−	−
	Person:Avail.	Compr.	GER	**The chief**$_{M}$ attacks soon and is martial, because [she/he] …	−	−	+
	Person:Avail.	Compr.	GER	**The chief**$_{M}$ is henceforth especially martial, because [she/he]…	−	+	+
QSCW (2012)	Availability	Compr.	CHI	…**Panzhen**$_{M}$ earned acclaim. [She/He] …	(+)	+	−
	Availability	Compr.	CHI	**Panzhen**$_{M}$ earned acclaim. In the new round of competition for outstanding individual [she/he]…	−	−	+
N (2014)	Availability	Fix	ENG	**The boy** thought that [she/he] …	+	−	+
	Availability	None	ENG	**The boy** thought that [she/he] …	+	−	−
	Availability	Compr.	ENG	**The boy** thought that [she/he] …	+	−	−
Xu (2015)	Availability	Compr.	CHI	**Grandpa** read Morning newspaper, [she/he] …	−	−	+
FCG (2018)	Availability	Compr.	ENG	[Ben$_{M}$/**Alice**$_{F}$] astonished Eric because she …	−	−	+
FFCG (2020)	Availability	Compr.	SPA	[Pablo$_{M}$/**Cataline**$_{F}$] met Rodrigo because she …	−	−	+

(9) a. The aunt heard that he . . .

x p
aunt(x)
hear(x,p)
p: z
male(z)
~~**z=x**~~
. . .

b. The aunt heard that she . . .

x p
aunt(x)
hear(x,p)
p: z
female(z)
z=x
. . .

Osterhout and Mobley (1995) report a P600 effect for the critical gender manipulation (9-a), compared to the gender-matching control condition (9-b). Since the (definitional) gender of the direct antecedent "The aunt$_F$" mismatches with the critical pronoun "he$_M$", and no other antecedents are available in the discourse, the use of the critical pronoun *he* has typically been described as a case of referential failure. Note, however, that the sentence does not constitute a strict syntactic or semantic anomaly: when imagining an unmentioned antecedent in the discourse, or embedding the sentence into a more elaborate (cataphoric) context with a suitable [male] referent (e.g., "The aunt heard that he won the lottery after her nephew bought his first ticket"), the apparent agreement violation can be resolved. Nevertheless, preliminary investigation of participants' individual tendencies in judging gender mismatch pronouns revealed that this effect was driven by the group of participants that typically judge gender mismatch sentences as unacceptable (Osterhout and Mobley, 1995). Indeed, according to referential RI theory, encountering the bare pronoun "he" (i.e., without a suitable antecedent) is predicted to result in increased integration difficulty, reflected in P600 amplitude, which may be due to either a revision/updating of the situation model that is being constructed, or a failure to achieve referential coherence.

Follow-up experiments that manipulated the availability of antecedents for pronoun resolution have replicated the P600 effect that was reported by Osterhout and Mobley (1995) (see, e.g., Nieuwland and van Berkum, 2006; Hammer et al., 2008; Xu, 2015; Fiorentino et al., 2018; Feroce et al., 2020). Some studies have also reported N400 effects for antecedent availability manipulations in languages other than English (e.g., German; Schmitt et al., 2002; Hammer et al., 2008, and Chinese; Xu, 2015). Furthermore, Nieuwland (2014) manipulated the task instructions regarding the processing of bare pronouns and observed a P600 effect when participants were explicitly instructed to fix cases of referential failure by imagining a novel referent for a mismatching pronoun. Interestingly, this P600 was not observed without explicit task instructions regarding bare pronouns. Nieuwland (2014) also reports a sustained Nref effect for bare pronouns under all task instructions, suggesting that participants in this particular experiment perceived bare pronouns as analogous to ambiguous pronouns. Indeed, while the introduction of a (gendered) definite noun phrase in principle does not prime away from introducing a novel entity into the discourse, differences in task, attention or grammatical constraints may lead to different referential expectations and thereby affect referential retrieval difficulty, as reflected in Nref amplitude.

3.3.2 Referential failure

Whenever the grammatical structure of the unfolding sentence constrains revision of the situation model under construction, mismatching referential expressions will lead to an increase in integration difficulty (P600) reflecting a failure to construct a coherent representation of the utterance meaning. This occurs, for instance, in cases where there is an agreement mismatch between a reflexive referential expression and its direct syntactic antecedent; for instance, Harris et al. (2000) presented participants with sentences such as "The pilot's mechanics brow-beat himself/themselves ...", where the singular "himself" constitutes a number agreement violation for the syntactic antecedent "mechanics". Harris et al. (2000) report a P600 effect, consistent with similar agreement manipulations (see Table 6). Hence, when a referential expression is used for which

Table 6: **0-Ref: Referential failure.** Lang. = language. Task abbreviations: Accep. = Acceptability judgments (typically every trial); Compr. = Comprehension questions (typically for a percentage of trials). Language abbreviations: DUT = Dutch; ENG = English; GER = German. Nref/N400/P600 modulations: "+" = significant effect reported; "−" = no significant effect reported; "(+)" = numeric difference.

Study	**Manipulation**	**Task**	**Lang.**	**Example: [Target/Control]**	**Nref**	**N400**	**P600**
O&M (1995)	Agreement:Number	Accep.	ENG	The hungry **guests** helped [himself/themselves] …	−	−	+
	Agreement:Gender	Accep.	ENG	The successful **woman** congratulated [himself/herself]...	−	−	+
OBM (1997)	Agreement:Gender	Accep.	ENG	The **woman** prepared [himself/herself] …	−	−	+
HWH (2000)	Agreement:Number	Comp.	ENG	The pilot's **mechanics** brow-beat [himself/themselves] …	−	−	+
LJHM (2006)	Agreement:Case	Compr.	GER	The **friend**$_{M}$ was invited and therefore they did count on [him$_{ACC}$/him$_{DAT}$] …	−	−	+
		Compr.	DUT	The **friend**$_{M}$ was invited and therefore they did count on [he$_{NOM}$/him$_{OBJ}$] …	−	−	+
CGO (2015)	Availability	Accep.	ENG	The **actress** prepared [himself/herself] …	−	−	+

there is no suitable and accessible antecedent and the syntactic structure does not allow for referent revision/introduction—i.e., 'true' cases of referential failure—this will lead to failure to achieve referential coherence, which manifests as increased integration difficulty. Critically, since both conditions in these contrasts concern reflexive pronouns, which therefore overlap in terms of their referential knowledge (i.e., coding for binding to a direct antecedent), no difference in referential retrieval difficulty is predicted to ensue for these contrasts.

4 Discussion

We have proposed an extension of Retrieval-Integration (RI) theory that incorporates referential processing. On the resultant referential RI theory, word meaning is assumed to consist of both conceptual knowledge and referential knowledge. The contents of the unfolding discourse/utterance representation in working memory may prime this knowledge in long-term memory, leading to facilitated retrieval in case the conceptual and referential knowledge associated with an incoming word is consistent with the state of the long-term memory system. The ease of the retrieval of conceptual knowledge is reflected in N400 amplitude, whereas the ease of the retrieval of referential knowledge is reflected in Nref amplitude. The actual establishment and/or revision of reference is attributed to the integrative processes underlying the P600. We have derived explicit predictions from referential RI theory in different referential constellations, and shown how these predictions are supported by empirical evidence. Finally, we have speculated how the referential retrieval hypothesis may extend to the processing of syntactic dependencies to explain structurally-induced LAN modulations in terms of retrieval.

Referential RI theory underlines the expectation-based nature of language comprehension. Expectation-based theories of comprehension, such as Surprisal theory (Hale, 2001; Levy, 2008), posit that upcoming linguistic input is conditionally predicted by the unfolding context. On referential RI theory, the automatic retrieval processes underlying the N400 and Nref are modulated by the degree to which word-associated conceptual and referential knowledge is pre-activated by the context. We assume this pre-activation to be a form of implicit prediction or expectation, consistent with the idea of 'readiness' of information in long-term memory (Gerrig and McKoon, 1998; Gerrig and O'Brien, 2005; see also Federmeier, 2022 for discussion). The pre-activation of conceptual and referential knowledge can be conceptualized as implicit predictions or expectations about incoming information, which critically may not always render the unfolding interpretation plausible (e.g., in role reversal anomalies, Hoeks et al., 2004; see Kuperberg, 2007; Bornkessel-Schlesewsky and Schlesewsky, 2008; Brouwer et al., 2012 for reviews). The unexpected nature of conceptually or referentially implausible continuations is instead reflected in the integrative processes underlying P600 amplitude. The expectation-based nature of these integrative processes—which may be modulated by task demands (Kolk et al., 2003) and attention (Schacht et al., 2014)—has led to the idea that the P600, like reading times, can be conceptualized as an index of 'comprehension-centric' surprisal (Venhuizen et al., 2019a; Brouwer et al., 2021b; Aurnhammer et al., 2021).[5] Indeed, consistent with this idea, the P600 has recently been shown to be graded for plausibility (Aurnhammer et al., 2023b).

The neurocomputational implementation of RI theory offers explicit, mechanistic instantiations of expectation-based retrieval and integration (Brouwer et al., 2017, 2021b). This neurocomputational model is effectively an extended simple recurrent neural network (SRN; Elman, 1990) that instantiates the *process* function defined in Equation 1. This model, which is schematically depicted in Figure 1 (bottom), consists of five layers of artificial neurons, and maps sentences on an incremental, word-by-word basis onto

[5] The conceptualization of the P600 as an index of 'comprehension-centric' surprisal suggests that the integrative processes underlying this component may be similar to those underlying the P300 (see Leckey and Federmeier, 2020; Sassenhagen et al., 2014, for discussion). Brouwer et al. (2021b) point out, however, that the P300 and P600 are differentially affected by task demands: While the elicitation of a P300 is strongly task-dependent, the P600 is modulated by task-demands, but no explicit task is required for its elicitation (Kolk et al., 2003).

  doi: 10.1037/rev0000530

a rich, "situation model"-representation of utterance meaning (see Venhuizen et al., 2019a,b, 2022), while taking the unfolding utterance context into account. Time in the model is discrete, and at each processing timestep t, the activation pattern in the **integration_context** layer represents the *utterance context* as established after processing the previous word at timestep $t-1$, and the activation pattern at the **input** layer represents a perceived *word form*. These activation patterns flow into a **retrieval** layer which maps the *word form* in the *utterance context* onto a representation of *word meaning* in the **retrieval_output** layer, thereby instantiating the *retrieve* function defined in Equation 2. Retrieved *word meaning* is then integrated into the *utterance context* in the **integration** layer producing an updated *utterance representation* at the **integration_output** layer, instantiating the *integration* function defined in Equation 3. Prior to processing the next word at $t+1$, the internal utterance representation represented in the activation pattern at the **integration** layer is copied to the **integration_context** layer in order to provide an updated *utterance context*.

Crucially, the model offers explicit linking hypotheses to N400 and P600 amplitude, and has been shown to account for key processing phenomena (Brouwer et al., 2017, 2021b; also see Crocker and Brouwer, 2023). Retrieval of word-associated conceptual information, as reflected in N400 amplitude, is estimated as the degree to which the activation pattern in the **retrieval** layer changes from one word to the next. The logic here is that if the conceptual knowledge associated with an incoming word is anticipated, this should be reflected in the state of the retrieval system of the model, and the activation pattern in the **retrieval** layer should show little change, as it is already consistent with the incoming information. If on the other hand, the conceptual knowledge associated with the incoming word is not anticipated, this will yield a large change in this activation pattern. Integration effort, in turn, as reflected in P600 amplitude, is estimated as the degree to which the activation pattern in the **integration** layer changes from one word to the next. Indeed, the more the meaning of an incoming word leads to a change in the unfolding utterance representation, the larger the change in this activation pattern, and vice versa. While the model has not been examined in light of referential processing, we predict referential knowledge, like conceptual knowledge, to modulate the activation pattern in the **retrieval** layer, such that different subsets of the neurons in this layer are more sensitive to referential versus conceptual aspects of word meaning, and that it is changes in activation in these different subsets that drive Nref versus N400 modulations, respectively (see also Crocker et al., 2010). Successful neurocomputational simulations of referential processing will be instrumental in turning theoretically proposed mechanisms into computationally explicit instantiations thereof.

The neurocomputational model directly aligns with the functional-anatomic mapping of RI theory onto a minimal cortical network (Brouwer and Hoeks, 2013; Brouwer et al., 2017). This mapping, which is schematically depicted in Figure 1 (top), is built around two cortical epicenters (or hubs): the posterior part of the left Middle Temporal Gyrus (lpMTG; Brodmann Area, BA, 21) serves as an epicenter for the retrieval of word meaning, and the left Inferior Frontal Gyrus (lIFG; BA 44/45/47) as an epicenter for the integration of retrieved word meaning into the unfolding utterance representation. It is the lIFG that mediates the utterance representation which includes the available discourse referents (see Nieuwland et al., 2007b, for fMRI evidence implying frontal regions in referential processing). The lpMTG and lIFG are wired together via white matter tracts in the dorsal pathway (dp) and ventral pathway (vp) (see Brouwer and Hoeks, 2013, for discussion on the directionality of these pathways). Depending on whether an incoming word is heard or read, an acoustic or orthographic word form is projected to the lpMTG from the auditory cortex (ac) or visual cortex (vc), respectively. The lpMTG then retrieves the conceptual knowledge associated with this word form, which is distributed across the association cortices, and the ease of this retrieval is reflected in the scalp-recorded N400 components. Retrieved word meaning is then projected to the lIFG, where it is integrated with the current utterance context into an

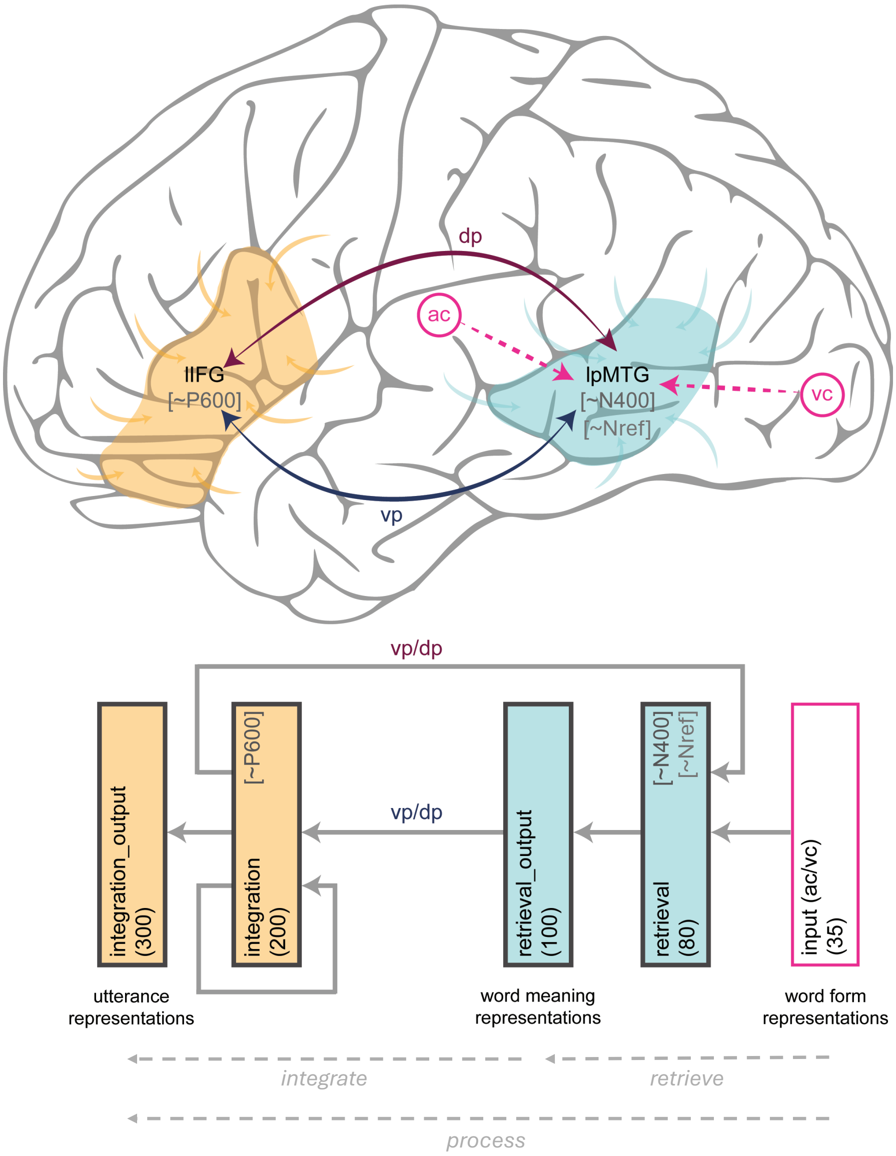

Figure 1: **Functional-anatomic Mapping of Retrieval-Integration theory (Top) and Neurocomputational Instantiation (Bottom).** Depending on the input modality, incoming words reach the lpMTG via either the auditory cortex (ac) or visual cortex (vc) (which correspond to the **input** layer in the model). The lpMTG then retrieves the conceptual and referential knowledge associated with an incoming word from the association cortices (**retrieval** → **retrieval_output**), while taking utterance context in the lIFG into account (**integration_context** → **retrieval**). Note that the schematic of the model uses a shorthand notation for the contextual input to the **retrieval** and **integration** layers (by omitting the **integration_context** layer). This retrieval process generates the N400/Nref component. The lIFG then receives this retrieved word meaning (**retrieval_output** → **integration**) and integrates it with the prior context (**integration_context** → **integration**), producing an updated utterance representation (**integration** → **integration_output**). This integration process generates the P600 component. dp = dorsal pathway; lIFG = left inferior frontal gyrus; lpMTG = left posterior middle temporal gyrus; vp = ventral pathway. Adapted with permission (CC BY-NC) from (Brouwer et al., 2017, p. 1345).

updated utterance representation, and it is these integrative processes that are reflected in the scalp-recorded P600 component. The cycle is then completed by feeding back the updated utterance representation in the lIFG to the lpMTG in order to pre-activate conceptual knowledge relevant to potential upcoming words.

Casting the retrieval of referential and conceptual knowledge as two parts of the same coin has implications for our conceptualization of the Nref and the N400. That is, these components may be generated by (partially) overlapping cortical generators in the lpMTG, and may therefore not be (easily) separable. If referential expressions carry predominantly referential knowledge (e.g., pronouns), retrieval effort may be reflected in a more frontally pronounced negativity, akin to the canonical Nref. For referential expressions that carry substantial conceptual knowledge (e.g., definite noun phrases introducing a novel entity), on the other hand, retrieval effort may be reflected in a more centroparietally distributed negativity, akin to the canonical N400. For referential expressions that carry both referential and conceptual knowledge, however, retrieval effort may be reflected in a scalp distribution that is anywhere on this fronto-to-centroparietal spectrum. This may in fact explain why studies investigating referential processing have reported Nref effects and N400 effects. To complicate matters further, these scalp-recorded negativities may be affected by spatiotemporal component overlap with the P600; that is, if the generators underlying the N400/Nref and P600 are active simultaneously, e.g., because retrieval and integration processes are partially cascaded or because the processing of a single word involves multiple RI cycles (Brouwer and Hoeks, 2013), the scalp-recorded signal will be the summation of the latent N400/Nref and the latent P600 components (Brouwer and Crocker, 2017; Brouwer et al., 2021a; Delogu et al., 2021).

In order to further understand the relationship of the Nref to the N400, as well as to account for spatiotemporal component overlap, extant and future data on referential processing could be (re-)analyzed using advanced data analysis techniques, such as regression-based ERP (rERP) estimation (Smith and Kutas, 2015; Brouwer et al., 2021a). In rERP analysis, linear regression is used to estimate the voltage measured for every subject at every electrode, time-stamp, and trial as a linear combination of properties of the stimulus presented at that trial. Such properties may, for instance, include the association of a target word to the context, its plausibility, and its suitability as a referential expression given the unfolding situation model. The degree to which these estimates approximate the observed signal can then be assessed by examining the average residuals by condition over time. The more these approximate zero, the better the fit of the rERP analysis. The coefficients of the resultant models are then informative about how the different properties combine, and possibly interact, in driving the observed signal over time. Given a well-controlled experimental design, this allows for identifying additive effects of stimulus properties (Aurnhammer et al., 2021), isolating and disentangling their different scalp distributions (Aurnhammer et al., 2023b), and critically, their potentially conflicting influence on the signal, which may lead to spatiotemporal component overlap (Brouwer et al., 2021a).

To summarize, we believe that by complementing our theoretical derivation of referential RI theory with explicit neurocomputational modeling and advanced data analysis techniques, we can further our understanding of the mechanisms and spatiotemporal dynamics of referential retrieval and integration in language comprehension. Proximately to further solidify referential RI theory as an integrated theory of the N400, Nref, and P600, but ultimately to also incorporate, for instance, structurally-induced LAN modulations in terms of retrieval.

5 Conclusion

Neurocognitive theories and models of language comprehension, informed by Event-Related Potentials (ERPs), typically focus on how semantic congruency and syntactic felicity affect the N400 and the P600, the two most salient components of the ERP signal, and do not explicate the mechanisms of context-dependent referential processing. To address this gap, we derived a mechanistic account of the processes underlying the Nref component, a frontal, sustained negativity that is sensitive to various types of referential uncertainty, by incorporating referential processing into Retrieval-Integration (RI) theory, an integrated theory of the N400 and the P600 in language comprehension. On RI theory, N400 amplitude reflects the contextualized retrieval of the conceptual knowledge associated with an incoming word from long-term memory, and P600 amplitude the integration of this word meaning into the unfolding utterance representation. The core premise of our referential extension to RI theory is that the meaning of a word in context is not only defined by its associated conceptual knowledge, but also by its associated referential knowledge in long-term memory. While N400 amplitude reflects the retrieval of conceptual knowledge, Nref amplitude reflects the retrieval of referential knowledge. Critically, referential RI theory does thus not implicate the Nref in the establishment of reference itself. The actual establishment and/or revision of reference is instead attributed to the integrative processes underlying the P600. Referential RI theory generates explicit predictions with regard to different referential constellations, including constellations in which there are multiple available referents (2-ref), a single available referent (1-ref), or no available referents (0-ref), and have we shown how these predictions are supported by the empirical evidence. Finally, we have have offered a starting point for further investigation using explicit neurocomputational modeling and advanced data analysis, and we have outlined a functional-neuroanatomic implementation of referential RI theory and discussed its implications for electrophysiology. Taken together, referential RI theory thus provides an integrated framework of the electrophysiology of the word-by-word construction, reorganization, and updating of (referential) meaning in online language comprehension.

References

Aurnhammer, C., Crocker, M. W., and Brouwer, H. (2023a). Single-trial neurodynamics reveal N400 and P600 coupling in language comprehension. *Cognitive Neurodynamics*, pages 1–17.

Aurnhammer, C., Delogu, F., Brouwer, H., and Crocker, M. W. (2023b). The P600 as a continuous index of integration effort. *Psychophysiology*, 60:e14302.

Aurnhammer, C., Delogu, F., Schulz, M., Brouwer, H., and Crocker, M. W. (2021). Retrieval (N400) and integration (P600) in expectation-based comprehension. *PLOS ONE*, 16(9):e0257430.

Barkley, C., Kluender, R., and Kutas, M. (2015). Referential processing in the human brain: An event-related potential (ERP) study. *Brain Research*, 1629:143–159.

Bornkessel-Schlesewsky, I. and Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain research reviews*, 59(1):55–73.

Bos, J. (2003). Implementing the binding and accommodation theory for anaphora resolution and presupposition projection. *Computational Linguistics*, 29(2):179–210.

Boudewyn, M. A., Long, D. L., Traxler, M. J., Lesh, T. A., Dave, S., Mangun, G. R., Carter, C. S., and Swaab, T. Y. (2015). Sensitivity to referential ambiguity in discourse: The role of attention, working memory, and verbal ability. *Journal of Cognitive Neuroscience*, 27(12):2309–2323.

Brouwer, H. and Crocker, M. W. (2017). On the proper treatment of the N400 and P600 in language comprehension. *Frontiers in Psychology*, 8:1327.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., and Hoeks, J. C. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, 41:1318–1352.

Brouwer, H., Delogu, F., and Crocker, M. W. (2021a). Splitting event-related potentials: Modeling latent components using regression-based waveform estimation. *European Journal of Neuroscience*, 53(4):974–995.

Brouwer, H., Delogu, F., Venhuizen, N. J., and Crocker, M. W. (2021b). Neurobehavioral correlates of surprisal in language comprehension: A neurocomputational model. *Frontiers in Psychology*, 12:615538.

Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446:127–143.

Brouwer, H. and Hoeks, J. C. (2013). A time and place for language comprehension: Mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience*, 7:758.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98(2):159–168.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *Neuroreport*, 18(17):1851–1854.

Bush, G. and Shin, L. M. (2006). The multi-source interference task: an fMRI task that reliably activates the cingulo-frontal-parietal cognitive/attention network. *Nature Protocols*, 1(1):308–313.

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., and 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.

Canal, P., Garnham, A., and Oakhill, J. (2015). Beyond gender stereotypes in language comprehension: Self sex-role descriptions affect the brain's potentials associated with agreement processing. *Frontiers in Psychology*, 6:1953.

Cheyette, S. J. and Plaut, D. C. (2017). Modeling the N400 ERP component as transient semantic over-activation within a neural network model of word comprehension. *Cognition*, 162:153–166.

Chow, W.-Y., Lau, E., Wang, S., and Phillips, C. (2018). Wait a second! Delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience*, 33(7):803–828.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., and Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12:769–786.

Coopmans, C. W. and Nieuwland, M. S. (2020). Dissociating activation and integration of discourse referents: Evidence from ERPs and oscillations. *Cortex*, 126:83–106.

Coulson, S., King, J. W., and Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and cognitive processes*, 13(1):21–58.

Crocker, M. W. and Brouwer, H. (2023). Computational psycholinguistics. In Sun, R., editor, *Cambridge Handbook of Computational Cognitive Sciences*, pages 890–920. Cambridge University Press, Cambridge, UK.

Crocker, M. W., Knoeferle, P., and Mayberry, M. R. (2010). Situated sentence processing: The coordinated interplay account and a neurobehavioral model. *Brain and Language*, 112(3):189–201.

Delogu, F., Brouwer, H., and Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition*, 135:103569.

Delogu, F., Brouwer, H., and Crocker, M. W. (2021). When components collide: Spatiotemporal overlap of the N400 and P600 in language comprehension. *Brain Research*, 1766:147514.

Eddine, S. N., Brothers, T., Wang, L., Spratling, M., and Kuperberg, G. R. (2024). A predictive coding model of the n400. *Cognition*, 246:105755.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2):179–211.

Federmeier, K. D. (2022). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology*, 59(1):e13940.

Feroce, N., Fiorentino, R., Covey, L., and Gabriele, A. (2020). Neural evidence for the processing of referential ambiguity and referential failure in Spanish. In Pascual y Cabo, D. and Elola, I., editors, *Current theoretical and applied perspectives on hispanic and lusophone linguistics*, volume 27, pages 153–174. John Benjamins Publishing Company.

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

Fitz, H. and Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, 111:15–52.

Frisby, S. L., Halai, A. D., Cox, C. R., Ralph, M. A. L., and Rogers, T. T. (2023). Decoding semantic representations in mind and brain. *Trends in Cognitive Sciences*, 27(3):258–281.

Garvey, C. and Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 5(3):459–464.

Gerrig, R. J. and McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, 26(2-3):67–86.

Gerrig, R. J. and O'Brien, E. J. (2005). The scope of memory-based processing. *Discourse Processes*, 39(2-3):225–242.

Geurts, B. (1999). *Presuppositions and pronouns*. Elsevier.

Hale, J. T. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, pages 1–8, Stroudsburg, PA. Association for Computational Linguistics.

Hammer, A., Jansma, B. M., Lamers, M., and Münte, T. F. (2008). Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. *Brain Research*, 1230:177–191.

Harris, T., Wexler, K., and Holcomb, P. (2000). An ERP investigation of binding and coreference. *Brain and Language*, 75(3):313–346.

Hoeks, J. C., Stowe, L. A., and Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19(1):59–73.

Hoeks, J. C. J. and Brouwer, H. (2014). Electrophysiological research on conversation and discourse processing. In Holtgraves, T. M., editor, *The Oxford Handbook of Language and Social Psychology*, pages 365–386. New York: Oxford University Press.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press, Cambridge, MA.

Kamp, H. (1981). A theory of truth and semantic representation. In Groenendijk, J. A. G., Janssen, T. M. V., and Stokhof, M. B. J., editors, *Formal Methods in the Study of Language, Proceedings of the Third Amsterdam Colloquium*, pages 277–322, Amsterdam. Mathematisch Centrum.

Kamp, H. and Reyle, U. (1993). *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and Discourse Representation Theory*. Kluwer, Dordrecht.

Kamp, H., van Genabith, J., and Reyle, U. (2011). Discourse Representation Theory. In Gabbay, D. M. and Guenthner, F., editors, *Handbook of Philosophical Logic*, volume 15, pages 125–394. Springer Netherlands.

Karimi, H., Swaab, T. Y., and Ferreira, F. (2018). Electrophysiological evidence for an independent effect of memory retrieval on referential processing. *Journal of Memory and Language*, 102:68–82.

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

Kolk, H. H., Chwilla, D. J., Van Herten, M., and Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85(1):1–36.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146:23–49.

Kutas, M. and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12):463–470.

Kutas, M. and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62:621–647.

Kutas, M. and Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205.

Kutas, M., van Petten, C., and Kluender, R. (2006). Psycholinguistics electrified II: 1994–2005. In Traxler, M. J. and Gernsbacher, M. A., editors, *Handbook of Psycholinguistics, 2nd Edition*, pages 659–724. Elsevier, New York.

Lamers, M. J., Jansma, B. M., Hammer, A., and 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(1):1–12.

Laszlo, S. and Plaut, D. C. (2012). A neurally plausible parallel distributed processing model of event-related potential word reading data. *Brain and Language*, 120(3):271–281.

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9(12):920–933.

Leckey, M. and Federmeier, K. D. (2020). The P3b and P600 (s): Positive contributions to language comprehension. *Psychophysiology*, 57(7):e13351.

Ledoux, K., Gordon, P. C., Camblin, C. C., and 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.

Li, J. and Ettinger, A. (2023). Heuristic interpretation as rational inference: A computational model of the N400 and P600 in language processing. *Cognition*, 233:105359.

Li, X. and Zhou, X. (2010). Who is ziji? ERP responses to the chinese reflexive pronoun during sentence comprehension. *Brain Research*, 1331:96–104.

Martin, A. E., Nieuwland, M. S., and Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage*, 59(2):1859–1869.

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

Marx, K., Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2007). Resolving complex anaphors. In Schwarz-Friesel, M., Consten, M., and Knees, M., editors, *Anaphors in Text: Cognitive, formal and applied approaches to anaphoric reference*, volume 86 of *Studies in Language Companion Series*, pages 259–278. John Benjamins Publishing.

McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37(4):547–559.

Meyer, L., Grigutsch, M., Schmuck, N., Gaston, P., and Friederici, A. D. (2015). Frontal-posterior theta oscillations reflect memory retrieval during sentence comprehension. *Cortex*, 71:205–218.

Molinaro, N., Barber, H. A., and Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, 47(8):908–930.

Muskens, R. (1996). Combining Montague semantics and discourse representation. *Linguistics and Philosophy*, 19(2):143–186.

Näätänen, R. and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24(4):375–425.

Nakamura, M., Momma, S., Sakai, H., and Phillips, C. (2024). Task and timing effects in argument role sensitivity: Evidence from production, EEG and computational modeling. Retrieved from https://www.colinphillips.net/wp-content/uploads/2024/02/nakamura2024.pdf.

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., Otten, M., and van Berkum, J. J. (2007a). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience*, 19(2):228–236.

Nieuwland, M. S., Petersson, K. M., and van Berkum, J. J. (2007b). On sense and reference: Examining the functional neuroanatomy of referential processing. *NeuroImage*, 37(3):993–1004.

Nieuwland, M. S. and van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, 1118(1):155–167.

Nieuwland, M. S. and van Berkum, J. J. (2008a). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language*, 106(2):119–131.

Nieuwland, M. S. and van Berkum, J. J. (2008b). The neurocognition of referential ambiguity in language comprehension. *Language and Linguistics Compass*, 2(4):603–630.

Osterhout, L., Bersick, M., and McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, 25(3):273–285.

Osterhout, L. and Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, 34(6):739–773.

Qiu, L., Swaab, T. Y., Chen, H.-C., and Wang, S. (2012). The role of gender information in pronoun resolution: Evidence from Chinese. *PLOS ONE*, 7(5):e36156.

Rabovsky, M., Hansen, S. S., and McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, 2(9):693–705.

Rabovsky, M. and McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, 132(1):68–89.

Sassenhagen, J., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late eeg positivity following linguistically deviant material is reaction time aligned. *Brain and language*, 137:29–39.

Schacht, A., Sommer, W., Shmuilovich, O., Martíenz, P. C., and Martín-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PloS one*, 9(3):e91226.

Schmitt, B. M., Lamers, M., and Münte, T. F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cognitive Brain Research*, 14(3):333–346.

Schumacher, P. B. (2011). The hepatitis called...: Electrophysiological evidence for enriched composition. In Steinbach, M. and Meibauer, J., editors, *Experimental Pragmatics/Semantics*, pages 199–219. John Benjamins.

Smith, N. J. and Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. the rERP framework. *Psychophysiology*, 52(2):157–168.

Steinhauer, K. and Drury, J. E. (2011). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, 120(2):135–162.

Streb, J., Hennighausen, E., and Rösler, F. (2004). Different anaphoric expressions are investigated by event-related brain potentials. *Journal of Psycholinguistic Research*, 33:175–201.

Swaab, T. Y., Camblin, C. C., and Gordon, P. C. (2004). Electrophysiological evidence for reversed lexical repetition effects in language processing. *Journal of Cognitive Neuroscience*, 16(5):715–726.

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A commentary on "grammatical agreement processing in reading: ERP findings and future directions" by Molinaro et al., 2014. *Cortex*, 66:149–155.

van Berkum, J. J., Brown, C. M., Hagoort, P., and Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology*, 40(2):235–248.

van Berkum, J. J., Koornneef, A. W., Otten, M., and Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, 1146:158–171.

van Berkum, J. J. A. (2009). The 'neuropragmatics' of simple utterance comprehension: An ERP review. In Sauerland, U. and Yatsushiro, K., editors, *Semantics and Pragmatics: From experiment to theory*, pages 276–316. Palgrave Macmillan, Basingstoke.

van Berkum, J. J. A., Brown, C. M., and Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41(2):147–182.

Van der Sandt, R. A. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, 9(4):333–377.

van Dijk, T. A. and Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Venhuizen, N. J., Bos, J., Hendriks, P., and Brouwer, H. (2018). Discourse semantics with information structure. *Journal of Semantics*, 35(1):127–169.

Venhuizen, N. J., Crocker, M. W., and Brouwer, H. (2019a). Expectation-based comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes*, 56(3):229–255.

Venhuizen, N. J., Crocker, M. W., and Brouwer, H. (2019b). Semantic entropy in language comprehension. *Entropy*, 21(12):1159.

Venhuizen, N. J., Hendriks, P., Crocker, M. W., and Brouwer, H. (2022). Distributional formal semantics. *Information and Computation*, 287:104763. Special Issue: Selected Papers from WoLLIC 2019, the 26th Workshop on Logic, Language, Information and Computation.

Wang, L. and Yang, Y. (2013). Integrating the meaning of person names into discourse context: an event-related potential study. *PLOS ONE*, 8(12):e83206.

Xu, X. (2015). The influence of information status on pronoun resolution in mandarin chinese: evidence from ERPs. *Frontiers in Psychology*, 6:873.

Zwaan, R. A. and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2):162–185.

