THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Concurrent use of animacy and event-knowledge during comprehension

Citation for published version:
Vega-Mendoza, M, Pickering, MJ & Nieuwland, MS 2021, 'Concurrent use of animacy and event-knowledge during comprehension: Evidence from event-related potentials', *Neuropsychologia*, vol. 152, 107724. https://doi.org/10.1016/j.neuropsychologia.2020.107724

Digital Object Identifier (DOI):
10.1016/j.neuropsychologia.2020.107724

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Neuropsychologia

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

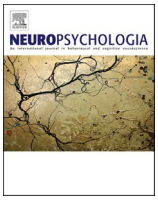
Download date: 17. Aug. 2021

Neuropsychologia 152 (2021) 107724

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia



Concurrent use of animacy and event-knowledge during comprehension: Evidence from event-related potentials

Mariana Vega-Mendoza [a,b,*], Martin J. Pickering [a], Mante S. Nieuwland [a,c,d,e]

[a] Department of Psychology, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, Scotland, UK
[b] Department of Business Administration, Technology and Social Sciences. Engineering Psychology, Luleå University of Technology, Luleå, Sweden
[c] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[d] Donders Institute for Cognition, Brain and Behaviour, Nijmegen, the Netherlands
[e] Heinrich-Heine-University, Düsseldorf, Germany

ARTICLE INFO

Keywords:
Semantic processing
Selection-restrictions
Real-world knowledge
Semantic relatedness
N400

ABSTRACT

In two ERP experiments, we investigated whether readers prioritize animacy over real-world event-knowledge during sentence comprehension. We used the paradigm of Paczynski and Kuperberg (2012), who argued that animacy is prioritized based on the observations that the 'related anomaly effect' (reduced N400s for context-related anomalous words compared to unrelated words) does not occur for animacy violations, and that animacy violations but not relatedness violations elicit P600 effects. Participants read passive sentences with plausible agents (e.g., *The prescription for the mental disorder was written by the psychiatrist*) or implausible agents that varied in animacy and semantic relatedness (*schizophrenic/guard/pill/fence*). In Experiment 1 (with a plausibility judgment task), plausible sentences elicited smaller N400s relative to all types of implausible sentences. Crucially, animate words elicited smaller N400s than inanimate words, and related words elicited smaller N400s than unrelated words, but Bayesian analysis revealed substantial evidence against an interaction between animacy and relatedness. Moreover, at the P600 time-window, we observed more positive ERPs for animate than inanimate words and for related than unrelated words at anterior regions. In Experiment 2 (without judgment task), we observed an N400 effect with animacy violations, but no other effects. Taken together, the results of our experiments fail to support a prioritized role of animacy information over real-world event-knowledge, but they support an interactive, constraint-based view on incremental semantic processing.

1. Introduction

People can process different forms of semantic information during sentence comprehension in an incremental and sometimes predictive fashion (e.g., Altmann and Mirković, 2009). There is evidence that one particular form of semantic information, namely animacy, plays an important role in incremental processing. ERP results suggest that comprehenders rapidly assess the animacy of incoming arguments and its compatibility with selectional restrictions imposed by preceding verbs (Nieuwland et al., 2013; Nieuwland & Van Berkum, 2005, 2006; Paczynski and Kuperberg, 2012; Szewczyk and Schriefers, 2013). However, it is an open question whether or not animacy selection restrictions are prioritized – that is, whether they are processed prior to and independently of other forms of semantic information, such as general real-world event-knowledge that allows people to evaluate sentence plausibility (e.g., De Swart and Van Bergen, 2019; McRae and Matsuki, 2009; Milburn et al., 2016; Nieuwland and Van Berkum, 2006; Paczynski and Kuperberg, 2012).

To investigate this question, we used the paradigm developed by Paczynski and Kuperberg (2012; henceforth, PK12), who investigated whether animacy-selection restrictions are prioritized over real-world event-knowledge. Their approach was to test whether a word's semantic relatedness to sentence context has a different impact on how people process violations of animacy compared to violations of real-world event-knowledge. Participants read passive sentences with post-verbal arguments that rendered the sentences plausible or implausible (Table 1), when asked to evaluate sentence plausibility as in PK12 (Experiment 1) or when not (Experiment 2). In the following sections,

* Corresponding author. Department of Business Administration, Technology and Social Sciences. Engineering Psychology, Luleå University of Technology, 97187, Sweden.
E-mail address: mariana.vega.mendoza@ltu.se (M. Vega-Mendoza).

https://doi.org/10.1016/j.neuropsychologia.2020.107724
Received 26 November 2019; Received in revised form 2 October 2020; Accepted 8 December 2020
Available online 5 January 2021
0028-3932/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

we briefly review relevant research on semantic aspects of sentence comprehension and then outline our predictions for the current study.

Research in psycholinguistics has shed light onto how semantic information is used during sentence comprehension (e.g., Marslen-Wilson et al., 1988; Marslen-Wilson and Tyler, 1980; Pickering and Traxler, 1998). It has been suggested, for example, that sentence comprehension occurs in an incremental fashion, in that the meaning of each new word is integrated into the unfolding meaning of the sentence as soon as it is encountered (e.g., Marslen-Wilson, 1973). Incremental processing enables comprehenders to rapidly realize when a word is semantically incompatible with prior context, a situation that leads to a processing cost relative to compatible words. Previous research has tried to elucidate whether certain types of semantic incompatibility between a word and its sentence context incur greater processing costs, or are more rapidly detected than others (e.g., Marslen-Wilson et al., 1988; Milburn et al., 2016; Staub et al., 2007; Warren and McConnell, 2007). Here, we consider the role of animacy constraints and event-knowledge and how semantic relatedness interacts with both forms of semantic information. Our introduction focuses on studies that used Event-Related Potentials (ERPs), the technique employed by PK12 and by us.

1.1. Effects of animacy

Most ERP research on animacy processing has used the N400 (Kutas and Hillyard, 1980), an ERP component strongly associated with semantic retrieval and integration processes (Kutas and Federmeier, 2011; Van Berkum, 2009). For example, words that meet the animacy requirements or preferences imposed by the preceding context elicit smaller N400 amplitudes than words that do not, suggesting a processing advantage for words that meet animacy restrictions or preferences (e.g., Nieuwland et al., 2013; Nieuwland and Van Berkum, 2005; 2006; Paczynski and Kuperberg, 2012). Sentence-initial animate nouns elicit smaller N400s than appropriately matched inanimate nouns (Weckerly and Kutas, 1999), a finding that is consistent with the tendency of English speakers to start sentences with animate entities (e.g., Branigan et al., 2008; see also Clark, 1965), and which therefore leads to a preference for sentence-initial animate nouns in listeners and readers. Similarly, Paczynski and Kuperberg (2011, Experiment 1) reported reduced N400s for inanimate patients (e.g., *At the homestead the farmer plowed the meadow …*) compared to equally plausible animate patients (e.g., *At the homestead the farmer penalized the labourer …*), consistent with the preference for patient roles to be filled by inanimate entities rather than animate entities.

In a study on Polish sentence comprehension, Szewczyk and Schriefers (2011) explored whether animacy violations are processed differently from other semantic violations. In Polish, masculine nouns are marked for animacy through inflectional morphology. They compared ERPs elicited by words that violated both animacy and semantic constraints of the context ('knitted an employee') or by words that violated only semantic constraints ('knitted a medicine'), versus congruent control words ('knitted a scarf'). The authors reported similar N400 effects to both types of violation, but larger P600 effects for the animacy violations. They concluded that animacy information is processed differently from other forms of semantic information.

In a later study, Szewczyk and Schriefers (2013) concluded that people anticipate the animacy of upcoming words in Polish. They used discourse contexts that were more predictive of either an animate or an inanimate direct object noun in the story-final sentence. In this sentence, pre-nominal adjectives (which in Polish are marked for animacy) that matched the presumed animacy of the upcoming direct object noun elicited smaller N400s than adjectives that did not match the presumed animacy. This is suggestive of animacy prediction, with the results being similar to those in studies investigating other forms of prediction such as gender (e.g., Fleur et al., 2020; Nieuwland, Arkhipova & Rodríguez-Gómez, 2020a; Van Berkum et al., 2005; Wicha et al., 2004; for a review, see Pickering and Gambi, 2018). However, because animacy information is available through the adjective marking, these findings may merely reflect the incremental use of animacy information as it becomes available, rather than actual pre-activation of animacy features.

These findings highlight the prominent role of animacy information in sentence processing. However, the extent to which animacy is

Table 1
Summary of the sentence types and sample sentence, adapted from Paczynski and Kuperberg (2012).

Sentence Type	Description of critical word (CW)	Example
		The prescription for the mental disorder was written by the … on paper.
(1) Control	Semantically related to the sentence context, meets expectations about the most likely agent to perform the action described by the verb (real-world knowledge), rendering the sentence plausible.	*psychiatrist*
Violation sentences:		
(2) Animate-Related	Semantically related, meets the animacy-selection restriction of the verb, but violates real-world knowledge expectations, thus rendering the sentence implausible.	*schizophrenic*
(3) Animate-Unrelated	Semantically unrelated, meets the animacy-selection restriction of the verb, but violates real-world knowledge expectations, thus rendering the sentence implausible.	*guard*
(4) Inanimate-Related	Semantically related, violates the animacy-selection restriction of the verb, thus rendering the sentence not only implausible but also impossible.	*pill*
(5) Inanimate-Unrelated	Semantically unrelated, violates the animacy-selection restriction of the verb, thus rendering the sentence not only implausible but also impossible.	*fence*

Note: We use these condition names because they refer to clear properties of the critical words and they are shorter and simpler than the names used in PK12. Our conditions (1)–(5) correspond to PK12's conditions (1) Control (2) Related Real-World Knowledge Violations (3) Unrelated Real-World Knowledge Violations (4) Related Animacy Selection Restriction Violations (5) Unrelated Animacy Selection Restriction Violations

prioritized over other forms of semantic information remains unclear, particularly when it is difficult to disentangle whether these effects could reflect processing of (im)plausibility (see also Milburn et al., 2016). Here, like PK12, we address this issue by investigating whether semantic relatedness has the same impact on how people process violations of animacy and those of real-world event-knowledge.

1.2. Semantic relatedness

Words that are semantically appropriate in their context are facilitated during sentence processing (for a review see Schumacher, 2012; for behavioral evidence see e.g., Ehrlich and Rayner, 1981; Schwanenflugel and Shoben, 1985). In ERP studies, implausible words that are related to the context elicit smaller N400s than unrelated, equally implausible words (Kutas and Federmeier, 2011). An early demonstration of this 'related anomaly' effect was reported by Federmeier and Kutas (1999). Their participants read a context that presumably led to an expectation for a certain word (e.g., *At the dinner party, I wondered why my mother wasn't eating her soup. Then I noticed that she didn't have a …*) followed by the predictable word (*spoon*), an implausible word from the same semantic category (*knife*) that shared semantic features with the predictable word, or an implausible word from a different semantic category (*bowl*). Although both implausible words elicited a larger N400 effect relative to the predictable word, the N400 effect was reduced for the words that shared more features with the predictable word, particularly in highly constrained contexts that led to strong expectations for a particular word.

Metusalem et al. (2012, Experiment 1) showed that such effects do not hinge on semantic category information but can arise from general, event-based relatedness. They presented participants with passages such as *A huge blizzard ripped through town last night. My kids ended up getting the day off from school. They spent the whole day outside building a big …* which then continued with a highly predicted word (*snowman*), an implausible word that was related to the described event (*jacket*), or an implausible word that was unrelated to the described event (*towel*). Again, event-related anomalous nouns elicited reduced N400s compared to unrelated anomalous nouns (see also Nieuwland, 2015).

These studies show that semantic processes initiated by an incoming word are affected by the semantic relationship between that word and the context (or words in the context; for discussion see Otten and Van Berkum, 2008). Such processing costs therefore do not always reflect the ease of integration of a word's meaning into the context (in which case those costs would presumably merely reflect plausibility, i.e., whether a word renders a sentence thus far plausible; see (Nieuwland et al., 2020b). The semantic processing costs associated with a word can also depend on the extent to which its semantic features (e.g., animacy) are pre-activated by the event described in the context (see also Ito et al., 2016). The related-anomaly paradigm, i.e., comparing processing costs of anomalous words that have or do not have a strong semantic relationship with the context, therefore offered a suitable approach for PK12 to investigate whether animacy information is prioritized over real-world event-knowledge. They investigated whether the related-anomaly effect differs for violations of animacy and of event-knowledge.

1.3. Evidence for priority of animacy information reported by Paczynski and Kuperberg (2012)

As we outline below, we used the paradigm used by PK12 to investigate whether animacy information is prioritized over real-world knowledge information. PK12 examined whether semantic relatedness facilitates the semantic processing of animacy violations in the same way as real-world event-knowledge violations. To do this they recorded ERPs while participants read sentences such as (1)–(5) in Table 1. The authors reported three main findings. Most crucially, they found N400 attenuation for related words compared to unrelated words, but only when words met the animacy requirements of the preceding verb (2 compared to 3, see Table 1), not for inanimate words that did not meet the animacy requirements of that verb (4 compared to 5). Secondly, they found no N400 effect for animate-related words compared to control words, despite the differences in plausibility between these conditions. Thirdly, they observed enhanced P600s for inanimate words (4 and 5) compared to control words (1) and animate words (i.e., a P600 effect of animacy), which they took as evidence that animacy violations incurred continued processing problems because they render a sentence impossible, not just improbable.

Based on these results, PK12 argued that animacy information is used before semantic relatedness comes into play, and that relatedness facilitates further processing only if the animacy constraint is met. We call this the *animacy-first hypothesis,* in which animacy is used as an early filter on further processing. PK12 considered two accounts for their findings. In the first, prediction-based account, readers use animacy restrictions along with context-information to pre-activate a subset of animate verb-argument candidates. The argument is that animacy-based predictions are supposedly easier than predictions from event-knowledge because animacy is a binary variable, and therefore have an earlier impact on processing. On this account, inanimate arguments were not pre-activated and therefore no attenuation by semantic relatedness occurred when these arguments were encountered (4 compared to 5).

In their second, integration-based account, animacy is prioritized over real-world event-knowledge during semantic integration. Our interpretation of this account is that comprehenders construct a compositional interpretation of a verb and an argument only if the verb's animacy requirements are compatible with the animacy of the argument. They therefore combine a verb that requires an animate agent with an animate agent, but not with an inanimate agent. If the agent is animate, they are then affected by the extent to which the actual agent is semantically related to plausible agents. In comprehending the sentences in Table 1, comprehenders would first determine whether the agent was animate or not. If not (i.e., sentences 4 and 5), they would not integrate the agent into the unfolding semantic representation for the sentence, and so relatedness would have no effect (thus explaining the lack of an N400 attenuation for 4 compared to 5). If it was animate (i.e., in sentences 1–3), they would integrate the agent into the sentential representation, and so relatedness would have an effect (thus explaining the attenuation for 2 compared to 3).

PK12's results therefore appear to support a functional distinction between violations of animacy selection restrictions (inanimate versus animate words) and those of real-world event-knowledge (unrelated versus related words). In addition, their results suggest that animacy information is used either to generate expectations about upcoming nouns, in a more constraining way than real-world knowledge, or as an initial selection restriction filter during sentence processing.

At face value, however, PK12's results appear inconsistent with other related anomaly effects. For example, in Nieuwland and Van Berkum (2005, 2006), animacy violations did not elicit an enhanced N400 compared to correct words when the animacy violations were strongly related to the context. Such results are consistent with a fully incremental, *constraint-based* account (e.g., MacDonald et al., 1994;

Marslen-Wilson et al., 1988; Trueswell et al., 1994; for a review, see MacDonald and Seidenberg, 2006). In such an account, animacy violations are more disruptive simply because they are less plausible than violations of real-world event-knowledge, but animacy information and real-world event-knowledge become available together. In such an account, event-relatedness has an impact regardless of animacy. For example, the sentence "The prescription for the mental disorder was written by the …" facilitates processing of related animate nouns such as 'doctor' or 'patient', but may also facilitate related inanimate nouns such as 'pill' or 'medicine', because their concepts are implicitly part of the described event (e.g., Ferretti et al., 2007; Gerrig and McKoon, 1998; Gerrig & O'Brien, 2005; Metusalem et al., 2012; Nieuwland, 2015). Also relevant, in the study by Metusalem et al. (2012), about a third of the related anomalies did not match the animacy constraints of the context. Animacy was not analyzed, therefore we do not know the contributions from different items, but the fact that they observed a related anomaly effect at all could be telling, because a substantial portion of animacy violation items would have made it rather hard to observe a related anomaly effect if those items did not contribute to the effect.

We also wish to highlight one important result of PK12 that appears to have been overlooked by the authors and that is inconsistent with their conclusions. The observed interaction between animacy and relatedness was entirely driven by differences between the related words (see Figs. 2 and 3 in PK12). In their follow-up tests after observing an interaction effect, PK12 only examined the effect of relatedness for animate and inanimate words separately, but they did not test the effect of animacy separately for related and unrelated words. If they had, they would have found that there was no animacy effect for unrelated words (the bar graphs show roughly identical estimates for animate-unrelated and inanimate-unrelated words). This result directly contradicts their conclusion that animacy selection restriction violations evoked an N400 effect. More generally, this result is inconsistent with a prioritized role of animacy, because under an animacy-first account animacy must have an initial effect on processing regardless of relatedness.

In sum, PK12's conclusions are inconsistent both with their own data and with other studies in the literature. Because no other studies to date have addressed the same issues as PK12, and because it remains unclear whether animacy and real-world event-knowledge are indeed processed independently (see also Milburn et al., 2016), we conducted a replication of PK12. If animacy information is indeed prioritized, as suggested by PK12 in their animacy-first hypothesis, our study might strengthen the evidence for a functional distinction between violations of animacy selection restrictions and violations of real-world event-knowledge. But on a more interactive, constraint-based view on incremental semantic processing (e.g., Filik and Leuthold, 2008; Marslen-Wilson et al., 1988; Matsuki et al., 2011; Metusalem et al., 2012; Nieuwland and Van Berkum, 2005, 2006), our study should find facilitation of related words regardless of animacy.

1.4. Predictions for the current study

We adopted the experimental paradigm of PK12 with some changes (see method section) to carry out a replication of their study. This allowed us to firstly assess the interplay between processing of different forms of semantic information (see Table 1) during sentence comprehension. Therefore, our predictions focused on the N400 component elicited by the critical words in the five conditions. According to the animacy-first hypothesis (suggested by PK12), we should observe an absent or smaller N400 effect of semantic relatedness for words that do not meet the animacy-restriction of the verb (inanimate agents) compared to those that do (animate agents). But according to the constraint-based hypothesis, comprehenders make use of all different sources of semantic information in a rapid and incremental fashion, irrespective of the type of information (Marslen-Wilson et al., 1988; see also Milburn et al., 2016). In that case, we expect to see the same effect of relatedness for animate and inanimate agents.

Although the focus of the present study is on the N400 component, we will also report the ERP effects in a later time window to examine possible post-N400 positive ERP effects (P600). PK12 and Szewczyk and Schriefers (2011) reported P600 effects for animacy violations but not for other violations, and concluded that these P600 effects reflected an attempt by readers to deal with a proposition that is semantically impossible (not just implausible). If that conclusion is correct, we expect to replicate the P600 effects for the inanimate conditions irrespective of semantic relatedness.[1]

2. Experiment 1

2.1. Method

2.1.1. Participants

Forty-eight[2] English native speaking right-handed participants who had normal or corrected-to-normal vision were paid to take part. All participants gave informed consent. After excluding participants who had too many artifacts (see results section for information on rates of discarded trials), the final sample consisted of 40 participants (27 women, mean age 23.4, $SD = 4.8$).

2.1.2. Materials

The stimuli consisted of English passive sentences. Each sentence was composed of an introductory context that introduced the patient followed by a verb requiring an animate agent noun (henceforth referred to as the *critical word* or *CW*), and two subsequent words. In the control condition (1), the CWs were animate agents that were plausible and semantically related to the preceding context. In the animate-related condition (2), the CWs were implausible, semantically related, animate agents. In the animate-unrelated condition (3), the CWs were implausible, semantically unrelated, animate agents. In the inanimate-related condition (4), the CWs were implausible, semantically related, inanimate agents. In the inanimate-unrelated condition (5), the CWs were implausible, semantically unrelated, inanimate agents. As the verb always required an animate agent, the inanimate conditions but not the animate conditions involved selection-restriction violations.

We developed our stimuli from the 120 items from PK12 together with 61 novel items. We created the novel items in order to have a large enough item set to present at least 30 trials per condition without presenting the same context sentence twice (as occurred in PK12). We reasoned that presenting each context sentence twice could cause participants to process each second context sentence and critical word in relation to the first encounter of that context sentence and critical word, possibly leading to strategic expectancies about the animacy of the upcoming CW.

We adapted the original materials and wrote the novel materials in the following ways. First, we rephrased some of the original sentences, changed both spelling and some words that are used only in American English to British English. Second, to quantify the degree of semantic relatedness between each critical word and its preceding context we followed the procedure used by PK12 to calculate Semantic Similarity Values (SSV) using Latent Semantic Analysis (LSA; http://lsa.colorado.edu; Landauer, 1998; Landauer et al., 1998). The SSV of each sentence

[1] Like PK12, we also computed ERPs for sentence-final words to investigate the downstream processing consequences of semantic anomalies. We report those findings as supplementary materials and offer a brief description in footnote 7.

[2] A previous version of this manuscript had a sample size of 22 participants, of which only 20 were ultimately included in the analysis. To increase statistical power and the number of observations, we tested an additional number of participants. The pre-registration of sample increase can be found at https://osf.io/se3pc and results from the initial sample in Version 1 of this paper at https://psyarxiv.com/2qbmp/.

Table 2
Characteristics of the experimental stimuli.

Sentence type	CW Length	CW Frequency	LSA values	Plausibility ratings
1.Control	7.52 (2.27)	0.88 (0.64)	0.23 (0.14)	6.7 (0.61)
2.Animate-Related	7.39 (1.96)	0.96 (0.63)	0.19 (0.12)	2.3 (1.06)
3.Animate-Unrelated	7.39 (1.96)	0.98 (0.62)	0.04 (0.03)	1.7 (0.78)
4.Inanimate-Related	6.83 (2.19)	0.96 (0.60)	0.19 (0.11)	1.3 (0.69)
5.Inanimate-Unrelated	6.80 (2.25)	0.98 (0.59)	0.04 (0.03)	1.0 (0.14)

Note. Mean values (Standard Deviations in parentheses).
CW = Critical word
CW length = number of letters
CW Frequency given as log-transformed word frequency per million words.
Pre-ratings plausibility scale: 1–7.

was obtained by averaging the LSA values between the critical word and the content words in its preceding context (http://psych.paczynski.net/lsa.html), but we imposed the further restriction that, for each item, neither of the unrelated conditions had a higher LSA value than either the control or the related conditions. Third, two words followed the critical noun, whereas PK12 used a varied number of sentence-final words. Fourth, we wrote sentences such that the implausible critical words could not be plausible when interpreted figuratively or could not easily form part of a plausible compound noun, and there was no sentence completion that would render the sentence plausible.

All sentences, truncated after the critical word of each condition, were rated for plausibility by 6 native speakers of English, who did not take part in the EEG study, using a 1–7 plausibility Likert scale. Because we planned to carry out an additional study with the same materials on native-Spanish speakers, we also obtained plausibility judgements from 6 native-Spanish speakers to fit both groups' ratings, which were also used to create our materials (see below). Participants were shown all five critical words following each item and rated each critical word for plausibility. To balance the number of plausible sentences containing animate and inanimate nouns, we also included 30 passive filler sentences, with three implausible animate critical words and two plausible inanimate critical words per sentence. The 1–7 scale for each critical word contained a question mark as an additional option, which participants were asked to circle if they did not know the critical word. Each rating participant saw one of two different lists, which only differed in the order of appearance of the same items. Based on these plausibility pre-ratings, we either removed or rephrased (and then re-tested for plausibility) any sentence that contained words not known by non-native speakers or sentences that did not match our plausibility expectations by either group, and thus selected a final set of 155 sentences (Table 2).

In the total set of items for the ERP experiment, 137 animate and 137 inanimate critical words appeared in two different violation conditions of different items (related in one item, unrelated in another item), but none of the critical words appeared in the same condition of different items. We used counterbalanced stimulus lists so that, in a given list, words in the control condition only appeared once, and at most 10 words in the violation conditions appeared twice in the experiment. We added 90 plausible sentence fillers (adapted from PK12) using verbs that were compatible with either animate or inanimate agents. Sixty fillers had inanimate agents and thirty had animate agents.

For the plausibility ratings analysis, a one-way ANOVA with all five sentence types showed a main effect of sentence type ($F_{4,770} = 1637.4$, $p < .001$). Follow up pairwise comparisons with Bonferroni correction showed that sentences in the control condition were rated higher than all 4 violation conditions (all $ps < .001$).[3] A 2-way ANOVA with animacy and relatedness as within subject factors showed that inanimate nouns were rated as more implausible than animate nouns ($F_{1,154} = 190.6$, $p < .001$), and unrelated nouns were rated as more implausible than related nouns ($F_{1,154} = 54.0$, $p < .001$). An interaction between animacy and relatedness ($F_{1,154} = 5.5$, $p = .020$) followed up with simple effects revealed that animate-related words were rated as more plausible than animate-unrelated words ($F_{1,154} = 31.7$, $p < .001$), and inanimate-related words were rated as more plausible than inanimate-unrelated words ($F_{1,154} = 30.8$, $p < .001$), but that the difference was greatest for the animate nouns.

2.1.3. Procedure

We constructed five lists, each containing one version of each item and a similar number of items per condition. The order of the 155 experimental sentences and 90 fillers was pseudorandomized so that there were never more than two trials in a row involving the same condition and never more than four experimental or control trials in a row, while keeping a fixed trial order per list.

Participants were randomly assigned to one of the five lists. They sat in front of the computer monitor in a dimly lit room. Sentences were presented word-by-word in black font on a white-background screen and the experiment was run using E-Prime 2.0®. Each trial started with a fixation point on the center of the screen for 450 ms, followed by a 200 ms blank screen. Following the procedure of PK12, each word appeared for 450 ms in the center of the screen followed by a 100 ms blank screen. At the end of the sentence there was a 750 ms blank screen and a question mark appeared on the center of the screen. Participants had been instructed to indicate as quickly as possible whether the sentence was plausible or implausible using one of two buttons on a control pad. The next trial started after participants had given their answer. The experiment began with 7 practice trials and was split into 6 equal-length blocks with optional short breaks between blocks. The experiment lasted about 45 min, together with about 30 min of EEG preparation time. Before the EEG recording, participants completed a language background questionnaire.

2.1.4. Electroencephalogram (EEG) recording and data processing

The EEG was recorded at a sampling rate of 512 Hz using a BioSemi ActiveTwo system (http://www.biosemi.com) with 64 EEG electrodes in an international 10–20 electrode configuration, two additional mastoid

[3] Because the plausibility ratings used a Likert-type scale, we also used non-parametric tests for these pairwise comparisons. Both parametric and non-parametric analyses for these comparisons yielded the same pattern of results.

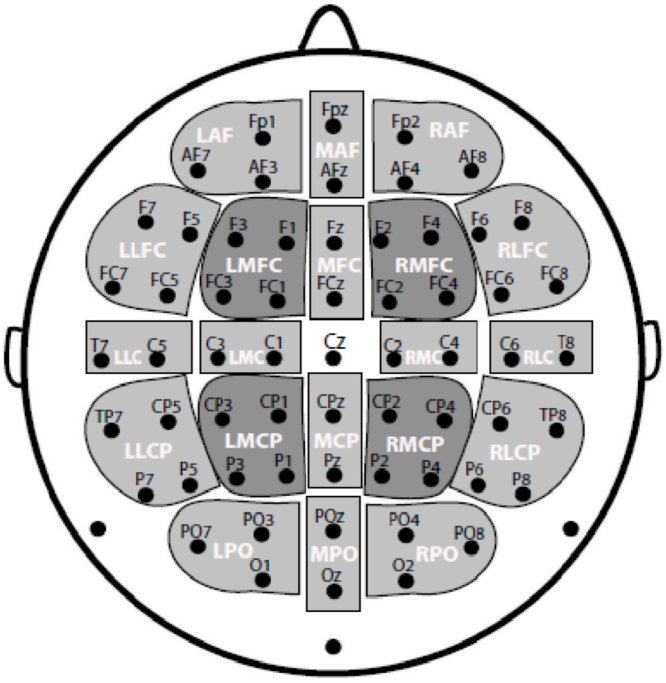


Fig. 1. Electrode configuration (black letters) and the Region of Interest clusters (white letters); medial ROIs used for analyses are shown in dark grey. The last one or two letters refer to the anterior/posterior dimension: AF = Anterior Frontal, FC = Frontocentral, C = Central, CP = Centroparietal, PO = Parieto-occipital. The first letter of 3-letter cluster-names and the first two letters of 4-letter cluster names refer to left-right dimension: L/R = Left/Right, LL/RL = Left/Right Lateral, LM/RM = Left/Right Medial (taken from Nieuwland, 2014).

electrodes and four EOG electrodes (left and right horizontal cantus, and above/below the right eye), referenced to the common mode sense (CMS; active electrode) and grounded to a passive electrode. The EEG was re-referenced to the average of the left and right mastoid electrode offline, filtered (0.05–20 Hz band-width filter[4] plus 50 Hz notch filter).

Data was segmented into epochs that started 100 ms before word onset, and that lasted until 1100 ms after CW onset. All epochs were corrected for ocular artifacts (Gratton and Coles correction; Gratton, Coles and Donchin, 1983), baseline-corrected using the 100 ms pre-CW time window, and then automatically screened for artifacts (allowed amplitude: minimal = −75 μV, maximal = 75 μV) before being entered into condition-averages per participant.

Cut-off was 16 artifact-free and correctly-responded CW epochs per condition. We excluded a total of 8 participants from further analysis. One participant had fewer than 16 correctly-responded trials in at least one condition, and 7 further participants had insufficient artifact-free trials. For the remaining 40 participants, averaged ERPs were computed over artifact-free trials for CWs per condition (Control: M = 24, SD = 3.7; Animate-Related: M = 22, SD = 3.6; Animate-Unrelated: M = 24, SD = 3.8, Inanimate-Related: M = 25, SD = 3.6; and Inanimate-Unrelated: M = 27, SD = 4.0).

2.1.5. ERP statistical analysis

We selected the same time windows for N400, P600, and sentence-final effects as PK12. We compared average ERP amplitude within each window per condition and clustered groups of electrodes into

[4] Upon reviewer request, we report an additional analysis with a 0.1–20 Hz filter setting in Appendix I.

Table 3
Mean percentages of condition-congruent plausibility judgments per condition from the ERP experiment.

Sentence Type	Plausible	Implausible
(1) Control	92 (5.5)	
(2) Animate-Related		83 (10.1)
(3) Animate-Unrelated		91 (7.6)
(4) Inanimate-Related		93 (5.5)
(5) Inanimate-Unrelated		99 (1.9)

Note. SD given in parentheses.

Regions-of-Interest (ROIs). These ROIs were organized into four groups (see Nieuwland, 2014), comparable to the 4-column approach reported by PK12: Lateral (LAF/RAF, LLFC/RLFC, LLCP/RLCP, LPO/RPO), Medial (LMFC/RMFC, LMCP/RMCP), Midline (MAF/MFC/MCP/MPO), and additional crossline (LLC/LMC/RMC/RLC) (Fig. 1). To simplify the presentation of our results, we only report analyses for the medial ROIs here. These medial ROIs are most important for observing the N400 and P600 components of interest, and correspond to the electrode regions where the effects observed by PK12 were maximal. Like PK12, we report two sets of analyses. The first set compares each violation condition directly to the control condition, through pairwise comparisons between each violation condition and the control condition using repeated measures Analyses of Variance (ANOVA) with sentence type as a 2-level factor (Sentence Type: control condition, violation condition) and the factors 2 (Hemisphere: left, right) by 2 (Anteriority: Frontal-Central, Central-Parietal) to test for scalp distribution effects.

The second set tests for interactions between animacy and relatedness in the four violation conditions, using 2 (Animacy: animate, inanimate) by 2 (Relatedness: related, unrelated) ANOVAs along with the same distribution factors as used in the first section. Where appropriate, Greenhouse–Geisser (Greenhouse and Geisser, 1959) p-values were applied and original F values are reported here. Only statistical results with $p < .1$ are reported. Interactions were followed up using simple effects analyses. We resolved interactions only when the involved factors were not part of higher-order significant interactions, which were resolved step-wise. Sentence-final effects are reported in supplementary materials (see also footnote 7).

2.2. Results

2.2.1. Behavioral results

2.2.1.1. Accuracy. The accuracy of the plausibility judgments is shown in Table 3. There was an overall effect of sentence type ($F_{4,156} = 31.1$, $p < .001$). Post-hoc pairwise comparisons with Bonferroni correction showed that animate-related sentences (2) were rated as less implausible than the remaining three violation conditions (3–5) (all $ps < .001$), whereas the inanimate-unrelated sentences (5) were rated as more implausible than the three other violation conditions (2–4) (all $ps < .001$). There were no differences in plausibility judgments between the

Table 4
Mean Reaction Times (ms) on accurate answers.

Sentence type	Reaction Time
(1) Control	459 (189)
(2) Animate-Related	497 (255)
(3) Animate-Unrelated	442 (200)
(4) Inanimate-Related	394 (145)
(5) Inanimate-Unrelated	350 (111)

Note: SD given in parentheses.

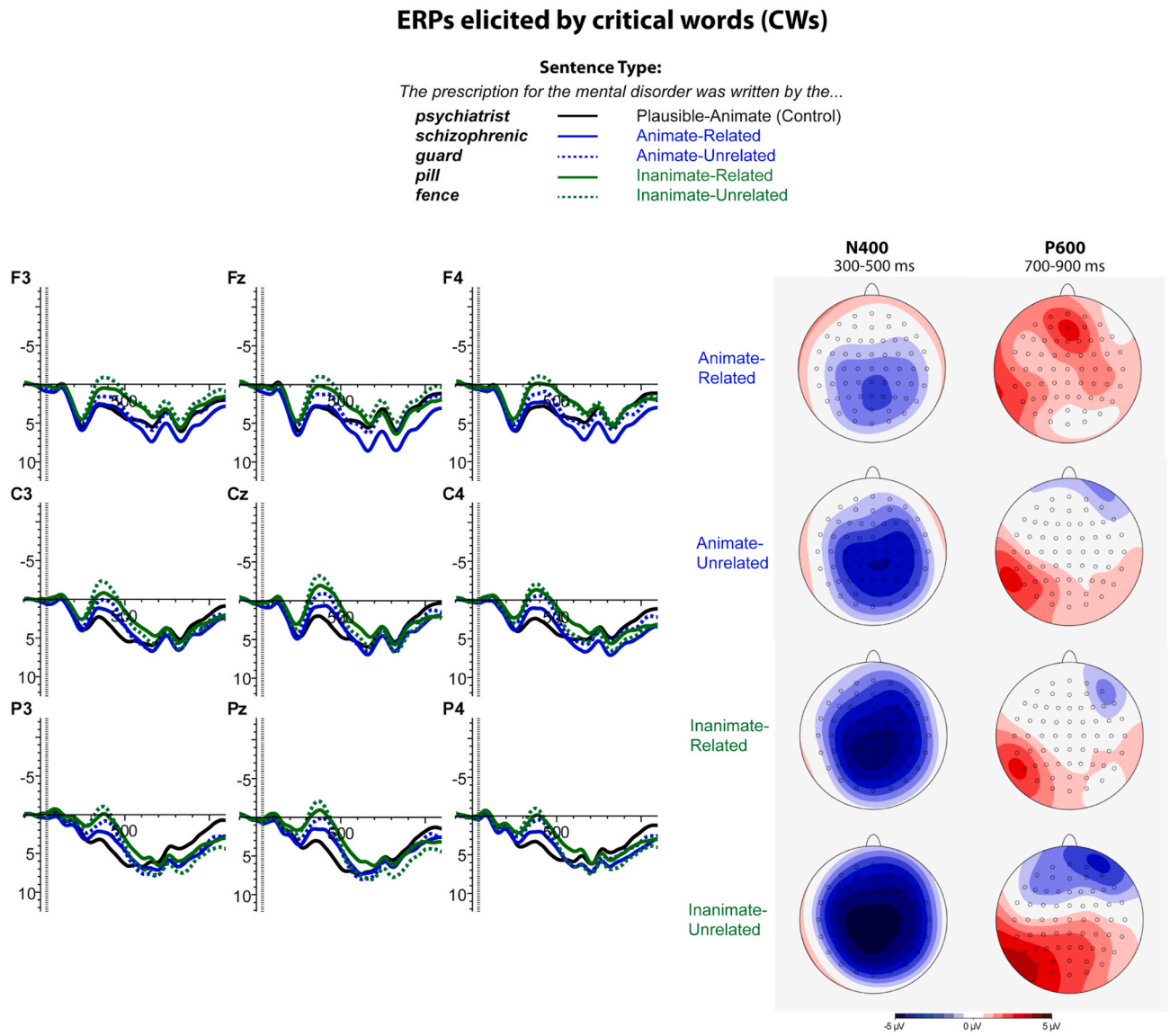


Fig. 2. ERPs elicited by critical words in Experiment 1. Topographical voltage maps show the difference with the control condition (i.e., anomaly minus control).

Table 5
F-value results from the pairwise ANOVAs contrasting each type of violation with the control condition at different time windows.

	N400 (300–500 ms)				P600 (700–900 ms)			
	Sent	Sent x Hem	Sent x AP	Sent x Hem x AP	Sent	Sent x Hem	Sent x AP	Sent x Hem x AP
(2) Animate-Related	4.3*	.20	8.5**	.003	5.0*	1.0	.64	1.8
(3) Animate-Unrelated	22.4***	2.7	6.3*	1.6	.43	.16	4.8*	7.0*
(4) Inanimate-Related	59.8***	1.2	3.5ˆ	5.1*	.30	.002	5.8*	1.2
(5) Inanimate-Unrelated	74.0***	.27	1.1	.72	1.5	.11	48.0***	8.5**

Notes. DF [1,39]. Sent: Sentence Type. AP: Anterior-Posterior distribution. Hem: Hemisphere distribution
ˆ$p < .1$
*$p < .05$
**$p < .01$
***$p < .001$

Table 6
F-value results from the 2 × 2 ANOVAs at the medial column.

	Main effect of Ani	Main effect of Rela	Ani x Rela	Ani x AP	Rela x AP	Ani x Hem x AP	Rela x Hem x AP	Ani x Rela x Hem	Ani x Rela x AP	Ani x Rela x Hem x AP
N400 (300–500 ms)	25.8***	8.7**	.60	2.3	.77	1.4	.08	1.9	.53	3.5^
P600 (700–900 ms)	1.1	.79	3.1^	19.0***	23.5***	.04	2.4	1.7	1.7	.20

Notes. DF [1,39]. AP: Anterior-Posterior distribution. Hem: Hemisphere distribution. Ani: Animacy. Rela: Relatedness.
^$p < .1$
*$p < .05$
**$p < .01$
***$p < .001$

animate-unrelated (3) and inanimate-related (4) conditions ($p = .852$).

2.2.1.2. Reaction time analysis. Reaction times (RTs) on correct answers for each participant were analyzed (Table 4), after removing values that fell outside 2 standard deviations from the mean per participant per condition. There was an overall effect of sentence type ($F_{4,156} = 11.2$, $p < .001$).

Post-hoc pairwise comparisons with Bonferroni correction showed that participants responded significantly slower to animate-related sentences (2) than inanimate-unrelated sentences (5) ($p < .001$) and inanimate-related sentences (4) ($p = .003$). Responses to the inanimate-unrelated sentences (5) were significantly faster than the other violation conditions (3–4) (all $ps \leq .004$). The animate-unrelated sentences (3) did not significantly differ from the animate-related (2) ($p = .422$) or inanimate-related (4) sentences ($p = .305$).

2.2.2. ERP results

2.2.2.1. N400 results

2.2.2.1.1. Pairwise (control vs. violation type) analyses. Grand-average ERPs per condition are shown in Fig. 2. Control words elicited significantly smaller (more positive) N400s ($M = 3.62$ μV, $SD = 3.57$) than animate-related words ($M = 2.35$ μV, $SD = 3.46$, $M_{diff} = 1.28$ μV, $\eta_p^2 = 0.099$), animate-unrelated ($M = 1.17$ μV, $SD = 3.46$, $M_{diff} = 2.45$ μV, $\eta_p^2 = 0.365$), inanimate-related ($M = 0.18$ μV, $SD = 3.22$, $M_{diff} = 3.44$ μV, $\eta_p^2 = 0.605$), and inanimate unrelated words ($M = -0.55$ μV, $SD = 3.41$, $M_{diff} = 4.18$ μV, $\eta_p^2 = 0.655$, Table 5). With regards to distributional effects, additional sentence type by anteriority interactions were found for the comparisons between control sentences and both animate conditions (Table 5). The N400s for control sentences were smaller than for animate-related sentences at posterior regions (control $M = 3.81$ μV, $SD = 3.75$; animate-related $M = 2.08$ μV, $SD = 3.66$, $M_{diff} = 1.73$ μV, $F_{1,39} = 6.7$, $p = .014$, $\eta_p^2 = 0.146$) but not at anterior regions (control $M = 3.44$ μV, $SD = 3.79$; animate-related $M = 2.61$ μV, $SD = 3.62$, $M_{diff} = 0.83$ μV, $F_{1,39} = 1.9$, $p = .176$, $\eta_p^2 = 0.046$). The N400s for control sentences were smaller than for the animate-unrelated sentences at both anterior and posterior regions, but the largest difference was observed at the posterior region (Anterior: animate-unrelated $M = 1.39$ μV, $SD = 3.92$, $M_{diff} = 2.05$ μV, $F_{1,39} = 13.6$, $p = .001$, $\eta_p^2 = 0.258$; Posterior: animate-unrelated $M = 0.95$ μV, $SD = 3.39$, $M_{diff} = 2.86$ μV, $F_{1,39} = 29.3$, $p < .001$, $\eta_p^2 = 0.429$). Finally, a three-way sentence by hemisphere by anteriority interaction for the comparison between control words and inanimate-related followed up at each quadrant separately, showed smaller N400s for control sentences at all quadrants, with the larger differences at the two posterior quadrants (LMFC: control $M = 3.43$ μV, $SD = 3.80$; inanimate-related $M = 0.54$ μV, $SD = 3.20$, $M_{diff} = 2.89$ μV, $F_{1,39} = 47.2$, $p < .001$, $\eta_p^2 = 0.548$; RMFC: control $M = 3.45$ μV, $SD = 3.86$; inanimate-related $M = 0.09$ μV, $SD = 3.32$, $M_{diff} = 3.35$ μV, $F_{1,39} = 59.9$, $p < .001$, $\eta_p^2 = 0.605$; LMCP: control $M = 3.90$ μV, $SD = 3.73$; inanimate-related $M = 0.09$ μV, $SD = 3.70$, $M_{diff} = 3.81$ μV, $F_{1,39} = 50.9$, $p < .001$, $\eta_p^2 = 0.566$; RMCP: control $M = 3.72$ μV, $SD = 3.93$; inanimate-related $M = -0.01$ μV, $SD = 3.70$, $M_{diff} = 3.73$ μV, $F_{1,39} = 45.4$, $p < .001$, $\eta_p^2 = 0.538$). There were no differences in distribution of N400 effects between control and inanimate-unrelated sentences as reflected by the lack of statistically significant interaction involving sentence type and the distribution factors anteriority and hemisphere.

2.2.2.1.2. Animacy by relatedness interaction analyses. Animate words elicited smaller N400s ($M = 1.76$ μV, $SD = 3.11$) than inanimate words ($M = -0.19$ μV, $SD = 3.10$, $M_{diff} = 1.95$ μV, $\eta_p^2 = 0.398$), and similarly, related words elicited smaller N400s ($M = 1.26$ μV, $SD = 3.05$) than unrelated words ($M = 0.31$ μV, $SD = 3.02$, $M_{diff} = 0.96$ μV, $\eta_p^2 = 0.183$), as indicated by main effects of animacy and relatedness (Table 6). There were no interactions involving both animacy and relatedness.[5] Additional analyses looking at the effect of animacy on relatedness showed that animate-related words ($M = 2.35$ μV, $SD = 3.46$) elicited significantly smaller N400s than inanimate-related words ($M = 0.18$ μV, $SD = 3.22$, $M_{diff} = 2.17$ μV, $F_{1,39} = 24.9$, $p < .001$, $\eta_p^2 = 0.390$) and animate-unrelated words ($M = 1.17$ μV, $SD = 3.46$) elicited significantly smaller N400s than inanimate-unrelated words ($M = -0.55$ μV, $SD = -0.55$, $M_{diff} = 1.72$ μV, $F_{1,39} = 11.1$, $p = .002$, $\eta_p^2 = 0.221$). Additional analyses looking at the effect of relatedness on animacy showed that animate-related words ($M = 2.35$ μV, $SD = 3.46$) elicited significantly smaller (more positive) N400s than animate-unrelated words ($M = 1.17$ μV, $SD = 3.46$, $M_{diff} = 1.18$ μV, $F_{1,39} = 6.0$, $p = .019$, $\eta_p^2 = 0.133$) and inanimate-related words ($M = 0.18$ μV, $SD = 3.22$) elicited marginally smaller N400s than inanimate-unrelated words ($M = -0.55$ μV, $SD = 3.41$, $M_{diff} = 0.73$ μV, $F_{1,39} = 3.8$, $p = .059$, $\eta_p^2 = 0.089$).

2.2.2.1.3. Summary of N400 results. Control words elicited significantly smaller N400s than all violation sentences. In the 2 × 2 analyses, animate nouns elicited smaller N400s than inanimate nouns, and likewise, semantically related nouns elicited smaller N400s than unrelated nouns. Importantly, no reliable interaction between animacy and semantic relatedness was found.

2.2.2.2. P600 results

2.2.2.2.1. Pairwise (control vs. violation type) analyses. Animate-related sentences elicited more positive ERPs than the control condition as shown by a main effect of sentence type (Table 5; control: $M = 4.82$ μV, $SD = 3.68$; animate-related $M = 6.29$ μV, $SD = 3.77$, $M_{diff} = -1.47$ μV, $\eta_p^2 = 0.114$). A sentence by anteriority interaction for inanimate-

[5] A marginally significant four-way interaction between animacy, relatedness, hemisphere and anteriority (Table 6) followed up with animacy by relatedness analyses on each quadrant only confirmed main effects of animacy and relatedness at all quadrants but no animacy by relatedness interactions at any quadrant.

related sentences did not show any effects of sentence type at either anterior ($F < 1$) or posterior regions ($F_{1,39} = 1.5$, $p = .226$, $\eta_p^2 = 0.037$). A three-way sentence by hemisphere by anteriority interaction was observed for the animate-unrelated and inanimate-unrelated sentences (Table 5). Follow up analyses at each quadrant did not show any further sentence effects for animate-unrelated words compared to control words (all $ps \geq .105$). Inanimate-unrelated words elicited more positive ERPs than the control condition at the left and right posterior ROIs (LMCP: Control $M = 4.74$ μV, $SD = 4.28$, Inanimate-Unrelated $M = 6.90$ μV, $SD = 3.59$, $M_{diff} = -2.16$ μV, $F_{1,39} = 12.8$, $p = .001$, $\eta_p^2 = 0.247$; RMCP: Control $M = 4.66$ μV, $SD = 4.03$, Inanimate-Unrelated $M = 6.43$ μV, $SD = 3.68$, $M_{diff} = -1.78$ μV, $F_{1,39} = 9.2$, $p = .004$, $\eta_p^2 = 0.190$). No sentence effects occurred at the LMFC or RMFC quadrants ($ps \geq .111$).

2.2.2.2.2. Animacy by relatedness interaction analyses. As Table 6 shows, we found no main effects of animacy (Animates $M = 5.76$ μV, $SD = 3.62$; Inanimates $M = 5.27$ μV, $SD = 2.98$, $M_{diff} = 0.49$ μV) or relatedness (Related $M = 5.68$ μV, $SD = 3.17$; Unrelated $M = 5.35$ μV, $SD = 3.20$, $M_{diff} = 0.33$ μV). However, a 2-way interaction between animacy and anteriority (Table 6) showed more positive ERPs for animates than inanimates at anterior (Animates $M = 5.73$ μV, $SD = 3.90$; Inanimates $M = 4.54$ μV, $SD = 2.80$, $M_{diff} = 1.20$ μV, $F_{1,39} = 5.1$, $p = .030$, $\eta_p^2 = 0.115$) but not at posterior regions (Animates $M = 5.78$ μV, $SD = 3.54$; Inanimates $M = 6.01$ μV, $SD = 3.45$, $M_{diff} = -0.22$ μV, $F < 1$). A two-way interaction of relatedness by anteriority (Table 6) revealed that related words elicited more positive ERPs than unrelated words at anterior regions (Related $M = 5.68$ μV, $SD = 3.01$, Unrelated $M = 4.59$ μV, $SD = 3.43$, $M_{diff} = 1.09$ μV, $F_{1,39} = 7.0$, $p = .012$, $\eta_p^2 = 0.151$), but not at posterior regions (Related $M = 5.68$ μV, $SD = 3.55$, Unrelated $M = 6.11$ μV, $SD = 3.27$, $M_{diff} = -0.44$ μV, $F_{1,39} = 1.3$, $p = .268$, $\eta_p^2 = 0.031$).

2.2.2.2.3. Summary of P600 results. In the pairwise comparisons, we observed that the animate-related condition elicited more positive ERPs than the control condition. In addition, the inanimate-unrelated condition elicited more positive ERPs than the control condition at the posterior ROIs. This result differs from that of PK12, who observed P600s to both animacy selection restriction violations compared to the control condition.

In the animacy by relatedness interaction analyses, we observed larger positivities to related than unrelated words at anterior regions. Likewise, we observed more positive ERPs to animate than inanimate words at anterior but not at posterior regions, in contrast to PK12, who observed larger P600s for animacy selection restriction than real-world knowledge violations with a posterior distribution.

2.2.2.3. Bayes factor analysis. To quantify the evidence for the null-hypothesis that animacy and relatedness did not show an interaction pattern on the N400, we performed a JZS Bayes factor repeated measures ANOVA (JASP Team, 2017; see also Morey and Rouder, 2015; Rouder et al., 2012) with default prior scales, using the data from Experiment 1 with the same factors as in the previous analyses. This analysis revealed that a model with main effects was preferred to the model with the interaction by a Bayes factor of 4.5. The data therefore provide substantial evidence (Jeffreys, 1961) against the hypothesis that animacy and relatedness interact in the N400 window.

2.3. Discussion of experiment 1

This experiment investigated whether animacy information is prioritized or privileged during sentence comprehension, as suggested by PK12 in their animacy-first hypothesis or whether comprehenders use animacy and other sources of semantic information concurrently, as predicted by a constrained-based hypothesis.

Pairwise comparisons showed that all violation conditions (animate-related, animate-unrelated, inanimate-related, and inanimate-unrelated) elicited enhanced N400s compared to the control condition. In addition, participants showed effects of animacy and of relatedness, evidenced by smaller N400s elicited by animate nouns compared to inanimate nouns and by smaller N400s elicited by related nouns compared to unrelated nouns.

Importantly, our N400 results did not reveal an interaction between animacy and semantic relatedness, in contrast with PK12, which suggests that participants recruited semantic information (both animacy and semantic relatedness) during sentence comprehension in an incremental way, that is, without giving animacy information priority. Unlike PK12, we did observe a statistically significant N400 effect for animate-related words compared to control words. We offer a fuller discussion of these findings in the General Discussion, along with a discussion of the differences between our study and PK12 that could be relevant to understand the discrepant findings.

The main finding in the post-N400 time window was that participants showed more positive ERPs for related words than unrelated words at anterior regions. Furthermore, unlike PK12, who observed P600 effects for animacy violations, our pattern went in the opposite direction, namely larger positivities for animate words compared to inanimate words with an anterior distribution. Due to the anterior distribution of our positive ERP effects, they could be qualified as a frontal post-N400 positivity (PNP) or anterior positivity instead. Therefore, we do not think that the observed post-N400 modulations relate to the detection of propositional impossibility (in contrast to PK12). Instead, these effects may arise from the task demands in the experiment, as the P600 deflection across conditions roughly patterned with the condition-difficulty that was evident from the behavioral responses, with greater P600 deflections for conditions with slower responses and lower accuracy. This raises the question to what extent the ERP findings result from the processes of interest (i.e., the processing of different forms of semantic knowledge associated with the experimental manipulations) or from the meta-linguistic, decision-related processes induced by the task.

To address the potential effects of task demands on the observed ERP results and the potential induced strategies during sentence evaluation, we performed a further experiment (Experiment 2). As we will outline below, recent ERP research on language comprehension has shown that task-demands can influence the results in both qualitative and quantitative ways, for example, with presence or absence of ERP effects depending on the task participants are asked to perform (e.g., Chwilla et al., 1995; Kolk et al., 2003; Nieuwland, 2014; cf. Nieuwland, 2015, 2016). In Experiment 2, we therefore addressed the potential impact of task demands by repeating Experiment 1 without the plausibility judgment task.

3. Experiment 2

In many ERP studies on language comprehension, participants are asked to perform a meta-linguistic judgment task such as judging sentence acceptability (for reviews, see e.g., Bornkessel-Schlesewsky and Schlesewsky, 2008; Kuperberg, 2007). This approach has several benefits. It enables researchers to avoid analysis of sentences that are not interpreted in the intended way (e.g., sentences in an implausible condition that are considered to be plausible). It also enables researchers to make a direct comparison between online ERP effects and the 'final interpretation' of a sentence as reflected in an overt evaluation (e.g., Nieuwland, 2016). Also, a judgment task might enhance engagement in or attention to the linguistic materials, which could boost observed effects compared to experiments without such a task.

However, the use of sentence acceptability judgment tasks also has some disadvantages. Asking participants to evaluate a sentence may alter how they understand the sentence. For example, the task may cause participants to pay more attention to specific parts of information contained in a sentence (e.g., animacy). In this respect, the task itself may introduce additional decision-based effects or strategies and thus limit on the generalizability of the results. Moreover, a particular problem arises when decision-related processes, such as those involved in task-based evaluation or response preparation, elicit ERP effects (e.g.,

decision-based P300 component) that spatially and/or temporally overlap with the ERP effects elicited by the experimental manipulation (for discussion, see Nieuwland, 2019; Roehm et al., 2007). The ERPs may then not reflect the effects of interest.

One way to understand the contributions of the task is to run the same experiment without a metalinguistic task (e.g., Kolk et al., 2003; Nieuwland, 2014; 2015; 2016; Roehm et al., 2007). Importantly, results of previous studies using this approach suggest that the N400 and the P600 ERP component can be affected by a judgment task. A well-known phenomenon is the depth-of-processing effect on the N400 observed in word-level experiments, in which the same linguistic materials elicit larger N400 effects if the task taps into their meaning rather than superficial aspects such as whether they contain a certain letter (e.g., Brown and Hagoort, 1993; West and Holcomb, 2000). Likewise, sentence plausibility judgments may boost N400 effects as they require participants to process the meaning of the sentences more deeply, although there are several studies where no clear boost is observed (Kolk et al., 2003; Nieuwland, 2015, 2016). The N400 effects that we observed for animacy and relatedness in Experiment 1 may thus be reduced in an experiment without the plausibility judgment task.

P600 effects are often enhanced by a judgment task, such as a grammaticality judgment task (e.g., Osterhout and Mobley, 1995). While syntactic anomalies typically elicit P600 effects with or without a judgment task, this may not or not consistently be the case for semantic anomalies that elicit P600 effects. A few ERP studies reported that P600 effects elicited by semantic anomalies are not only boosted by an acceptability/plausibility judgment task but are not elicited in absence of such a task. For instance, Kolk et al. (2003) showed a biphasic N400–P600 effect in Dutch sentences with selection restriction violations (e.g., *The tree that in the park played/stood …)* with an acceptability judgment task, but only an N400 and no P600 without that task. Similarly, Schacht et al. (2014) showed absence of P600s and only an anterior negativity to Spanish semantically-violated sentences (e.g., *El sentimiento peludo/profundo emociona;* English*: The hairy/deep feeling moves*) when using a word probe task, whereas another study reported an N400–P600 effect on the same stimuli using acceptability judgements (Martin-Loeches, Nigbur, Casado, Hohlfeld and Sommer, 2006). This pattern of results led the authors to conclude that the semantic processes associated with P600 effects are not automatic, but task- or attention-driven (see also Kuperberg, 2007; Bornkessel-Schlesewsky and Schlesewsky, 2008).

At present it is unclear whether boosting of P600s effects by a task means the linguistic processes reflected in P600 are enhanced or that there is spatial-temporal overlap with decision-based positive components such as the P300 (see also Osterhout, 1999; Sassenhagen and Bornkessel-Schlesewsky, 2015; Sassenhagen et al., 2014). Regardless, based on previous results we expected that repeating Experiment 1 without the judgment task would lead to a reduction or disappearance of the observed P600 effects. In order to reduce any potential task-induced effects, we removed the plausibility sentence evaluation task and instead had participants read for comprehension.

In Experiment 2, participants did not perform an acceptability judgment task, but they answered occasional comprehension questions that were independent of the manipulations of interest. These questions were included to maintain participants' attention to the sentences. We hypothesized that we would observe smaller N400 effects than we observed in Experiment 1, and smaller or even absent P600 effects. An important caveat, however, is that in contrast to Experiment 1 in which participants evaluated each sentence for plausibility, trials in this experiment were not selected based on whether a participant found the sentence plausible or implausible, diluting the basic effects of plausibility. Therefore smaller effects are to be expected simply because the analysis includes implausible sentences which - had there been a task - participants may judge to be acceptable (and vice versa for plausible sentences).

3.1. Method

3.1.1. Participants

Twenty-one native speakers of English who had normal or corrected-to- normal vision took part in this study after signing informed consent. Participants received monetary compensation for their participation and they did not take part in the study described in Experiment 1. After removing participants who had too many artifacts or incomplete data (see results section, for details), the final sample consisted of 17 participants (12 women, mean age 21.9, $SD = 5.2$).

3.1.2. Materials and procedure

The sentence stimuli and experimental procedures were identical to those in Experiment 1, with the important exception of the task instruction. Instead of evaluating sentence plausibility, participants answered 80 yes/no comprehension questions. These questions were only added to ensure attention throughout the experiment, and probed knowledge about the sentences that was independent of sentence plausibility (e.g., *On the lake, the boat race was won by the (1) rower, (2) sunbather, (3) sheriff, (4) paddle, (5) revolver …. Question: Was the race at the lake?*) Of these questions, 50 were for experimental items and 30 for fillers, with an even distribution of yes/no correct answers. Each trial began immediately after participants pressed any key on the keypad on seeing the fixation point at beginning of each sentence.

3.1.3. Electroencephalogram (EEG) recording and data processing

The EEG recording parameters and data processing procedures were identical to those in Experiment 1. Based on the cut-off of 16 artifact-free CW epochs per condition, we excluded data from 3 participants due to excessive artifacts, and data from 1 participant due to incomplete data. For the remaining 17 participants, averaged ERPs were computed over artifact-free trials for CWs per condition. The average number of trials per condition on all participants in the final sample were as follows: Control: $M = 25$, $SD = 4.0$; Animate-Related: $M = 24$, $SD = 4.4$; Animate-Unrelated: $M = 25$, $SD = 4.0$, Inanimate-Related: $M = 24$, $SD = 4.9$; and Inanimate-Unrelated: $M = 25$, $SD = 3.8$.

3.1.4. ERP statistical analysis

Electrode-clustering and statistical analysis was identical to those reported in Experiment 1. Analyses were performed in 2 time windows: 300–500 ms and 700–900 ms relative to onset of the critical word. Additional analyses at the 300–500 ms relative to onset of the sentence-final word are reported in supplementary materials (see also footnote 6). As in Experiment 1, the first section compares each violation condition directly to the control condition, through pairwise comparisons, and the second section tests the effects of animacy and relatedness in the four violation conditions.

3.2. Results

3.2.1. Behavioural results

Mean accuracy percentage of condition-congruent responses to the yes/no comprehension questions was 96 ($SD = 3.8$) for Experimental items, and 95 ($SD = 4.7$) for Filler items.

3.2.2. ERP results

Visual inspection of the data indicates that the control condition elicited smaller N400s in the 300–500 ms time window than implausible conditions (see Fig. 3). These N400 effects were widely distributed and visible at most channels. For most conditions, the observed N400 effects extended, at least visually, beyond 500 ms and even into the later 700–900 ms analysis window (see footnote 6).

3.2.2.1. N400 results

3.2.2.1.1. Pairwise (control vs. violation type) analyses. The control

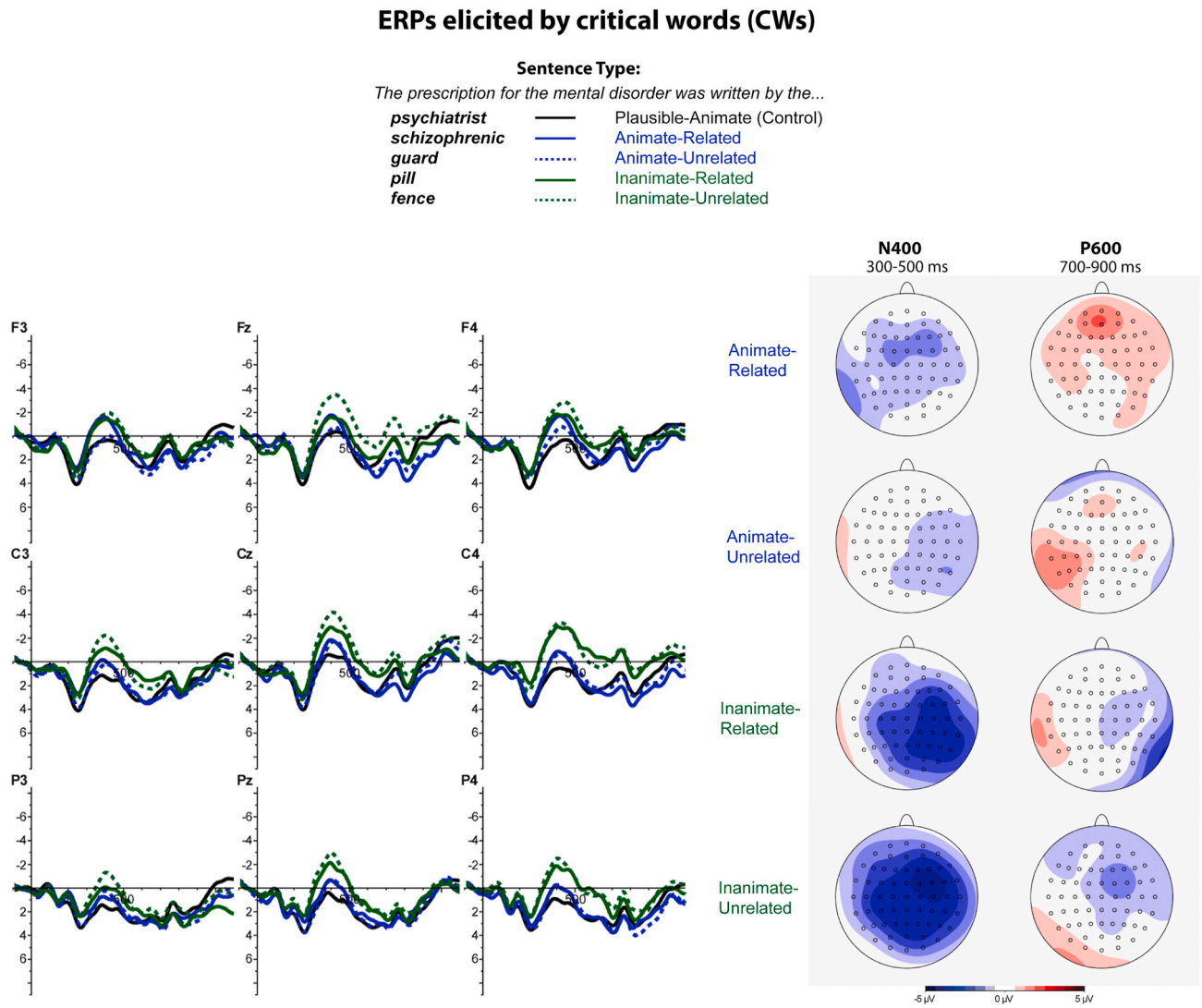


Fig. 3. ERPs elicited by critical words in Experiment 2 (no plausibility judgement task). Topographical voltage maps show the difference with the control condition (i.e., anomaly minus control).

condition elicited smaller N400s (M = 0.85 µV, SD = 2.37) than both inanimate related (M = −1.24 µV, SD = 2.62, M_{diff} = 2.09 µV, η_p^2 = 0.390) and inanimate unrelated (M = −2.00 µV, SD = 3.72, M_{diff} = 2.86 µV, η_p^2 = 0.495) conditions (Table 7). An interaction between sentence type and hemisphere for inanimate-related words showed smaller N400s for control words than inanimate-related words at both hemispheres, but the largest difference was observed at the right hemisphere (Left: Control M = 0.94 µV, SD = 2.55, Inanimate-related M = −0.76 µV, SD = 2.84, M_{diff} = 1.70 µV, $F_{1,16}$ = 5.6, p = .031, η_p^2 = 0.260; Right: Control M = 0.77 µV, SD = 2.34, Inanimate-related M = −1.71 µV, SD = 2.55, M_{diff} = 2.49 µV, $F_{1,16}$ = 16.4, p = .001, η_p^2 = 0.506). For inanimate-unrelated words, a three-way interaction between sentence type, hemisphere, and anteriority (Table 7) was followed up on each of the four quadrants, showing that control words elicited smaller N400s than inanimate-unrelated words at all quadrants (LMFC: Control M = 0.51 µV, SD = 2.74, Inanimate-Unrelated M = −1.90 µV, SD = 3.89, M_{diff} = 2.41 µV, $F_{1,16}$ = 9.8, p = .007, η_p^2 = 0.379; RMFC: Control M = 0.70 µV, SD = 2.69, Inanimate-Unrelated M = −2.54 µV, SD = 3.86, M_{diff} = 3.24 µV, $F_{1,16}$ = 14.2, p = .002, η_p^2 = 0.469; LMCP: Control M = 1.37 µV, SD = 2.71, Inanimate-Unrelated M = −1.49 µV, SD = 3.87, M_{diff} = 2.86 µV, $F_{1,16}$ = 15.4, p = .001, η_p^2 = 0.491; RMCP: Control M = 0.84 µV, SD = 2.44, Inanimate-Unrelated M = −2.08 µV, SD = 4.31, M_{diff} = 2.92 µV, $F_{1,16}$ = 11.5, p = .004, η_p^2 = 0.419), but the largest difference was observed at the right anterior (RMFC) ROI. No main effect of sentence type or interaction effect was found for the animate-related and animate-unrelated conditions.

3.2.2.1.2. Animacy by relatedness interaction analyses. Animate nouns elicited smaller N400s (M = −0.07 µV, SD = 2.61) than inanimate nouns (M = −1.62 µV, SD = 2.79, M_{diff} = 1.55 µV, η_p^2 = 0.374; Table 8). No main effect of relatedness or animacy by relatedness interactions were found.

3.2.2.1.3. Summary of N400 results[6]. Results at this time window showed smaller N400s to the control condition compared to the inanimate-related and inanimate-unrelated conditions. In the animacy by relatedness analyses, animate nouns elicited smaller N400s than inanimate nouns, but no effects of relatedness or animacy by relatedness interactions were obtained.

[6] Additional animacy by relatedness interaction analyses of the 500–700 ms time window confirmed that animate words elicited significantly smaller extended-N400s (M = 2.29 µV, SD = 2.73) than inanimate words (M = 0.89 µV, SD = 2.74; M_{diff} = 1.40 µV, $F_{1,16}$ = 7.1, p = .017, η_p^2 = 0.307). No effects of relatedness or interactions were found.

Table 7
F-value results from the pairwise ANOVAs contrasting each type of violation with the. control condition at different time windows.

	N400 (300–500 ms)				P600 (700–900 ms)			
	Sent	Sent x Hem	Sent x AP	Sent x Hem x AP	Sent	Sent x Hem	Sent x AP	Sent x Hem x AP
(2) Animate-Related	2.8	0.29	0.48	1.1	0.81	0.27	0.46	0.90
(3) Animate-Unrelated	2.5	3.3ˆ	0.60	0.05	0.35	1.8	0.39	0.35
(4) Inanimate-Related	10.2**	8.3*	0.51	0.50	0.14	4.3ˆ	0.01	0.02
(5) Inanimate-Unrelated	15.7**	1.7	0.01	5.6*	0.66	1.1	0.42	0.75

Notes. DF [1,16]. Sent: Sentence Type. AP: Anterior-Posterior distribution. Hem: Hemisphere distribution
ˆ$p < .1$
*$p < .05$
**$p < .01$
***$p < .001$

Table 8
F-value results from the 2 × 2 ANOVAs at the medial column.

	Main effect of Ani	Main effect of Rela	Ani x Rela	Ani x Hem	Rela x Hem	Ani x Rela x Hem	Ani x Rela x AP	Ani x Rela x Hem x AP
N400 (300–500 ms)	9.6**	0.14	1.2	1.4	0.07	2.8	1.3	1.7
P600 (700–900 ms)	4.0ˆ	0.36	0.02	0.70	0.44	1.1	0.37	0.27

Notes. DF [1,16]. AP: Anterior-Posterior distribution. Hem: Hemisphere distribution. Ani: Animacy. Rela: Relatedness.
ˆ$p < .1$
*$p < .05$
**$p < .01$
***$p < .001$

3.2.2.2. P600 results. In the pairwise comparisons, no sentence effects were observed. No effects of or interactions between animacy and relatedness were observed either (Table 8).

3.3. Discussion of experiment 2

We repeated Experiment 1 with only one change in the procedure, namely that participants did not evaluate sentence plausibility, but rather, answered occasional yes/no comprehension questions. We observed reliable effects of animacy, but not of relatedness. Inanimate nouns elicited N400 effects compared to the control nouns and compared to animate nouns, whereas N400s for the animate nouns did not differ from the control nouns. The N400 effects of animacy extended beyond the 300–500 ms time window. We did not observe reliable effects of semantic relatedness or interactions between animacy and relatedness.

4. General discussion

In two ERP experiments, we investigated whether animacy information is prioritized over real-world event-knowledge during sentence comprehension. We used the paradigm employed by Paczynski and Kuperberg (2012, PK12), who concluded that animacy information is prioritized (i.e., used as a filter on further semantic processing), such that retrieval of lexical-semantic information proceeds differently for words that do or do not violate animacy selection restrictions. Participants read passive sentences in English, with plausible (control) agents (e.g., *The prescription for the mental disorder was written by the psychiatrist),* or implausible agents that were either animate or inanimate and either semantically related or unrelated to the sentential context (*schizophrenic/guard/pill/fence*). We also examined the role of task-demands by performing the same experiment twice, once with the plausibility judgment task used by PK12 (Experiment 1) and once without (Experiment 2).

Before discussing our own results, we briefly summarize the three main results from PK12 to remind the reader of their evidence in support of the animate-first account. (1) The crucial finding was an interaction wherein related words elicited smaller N400s than unrelated words when these words were animate, but not when they were inanimate. (2) Animate-related words did not elicit a reliable N400 effect compared to control words. (3) Inanimate words elicited a P600 effect compared to control words, whereas animate words did not.

In our own two experiments, we found a very different pattern of results from PK12 on each of these three issues, and our results yielded no such evidence for the priority of animacy. We will discuss our results in comparison to the three core findings of PK12 listed above.

4.1. N400 results

Most importantly, in neither experiment did we observe an interaction between animacy and semantic relatedness on N400 activity. In Experiment 1, we observed two main effects on the N400: Animate nouns elicited smaller N400s than inanimate nouns, and related nouns elicited smaller N400s than unrelated nouns. The effect of animacy was stronger than that of relatedness. In Experiment 2, we observed only the main effect of animacy. These results are not compatible with the animacy-first account proposed by PK12, but lend support to an interactive, constraint-based account in which animacy influences processing at the same stage as real-world knowledge (MacDonald et al., 1994; Marslen-Wilson et al., 1988; Trueswell et al., 1994). In addition, these patterns could reflect the facilitation of semantic retrieval, or easier integration of the animate nouns and related nouns compared to inanimate nouns and unrelated nouns. These results are consistent with those of other studies that reported reduced N400s for words that match the sentence context in terms of animacy (e.g., Nieuwland and Van Berkum, 2006; Szewczyk and Schriefers, 2013), or for words that are semantically related to the discourse context compared to unrelated words (e.g., Ito et al., 2016; Ito et al., 2017; Metusalem et al., 2012; Nieuwland, 2015).

One way to interpret our results is in terms of semantic prediction (e.g., Federmeier and Kutas, 1999; Ito et al., 2016). Coarse-grained semantic predictions work through the activation of semantic information that is associated with the described event, which facilitates semantic retrieval processes elicited by the noun (e.g., Van Petten and Luka, 2012). Semantic prediction is thought to underlie the 'related-anomaly effect', the finding that implausible words that are related to the described event of the predicted word lead to smaller N400s than unrelated words (e.g., Ito et al., 2016; Metusalem et al., 2012; Nieuwland,

2015). The selection restrictions available in each sentence context in our experiment could lead to pre-activation of animacy, whereas the described event as a whole may pre-activate general, event-related information. For example, after reading "*the prescription was written by the*", people may expect the next upcoming word to be animate (e.g., Szewczyk and Schriefers, 2013). At the same time, activation of semantically related concepts arises from event-knowledge, for example the activation of '*pill*' due to knowledge about prescription medicine, even if this word constitutes an anomalous continuation at that point in the sentence due to its animacy. Thus, both sources of information may influence knowledge activation simultaneously, and ultimately have the same facilitatory effect on the semantic retrieval processes elicited by the noun.

An alternative explanation is that the results reflect ease of integration (e.g., Brown and Hagoort, 1993; for discussion of ease of integration accounts, see Ito et al., 2016; Nieuwland et al., 2020b; Pickering and Gambi, 2018). The facilitation of animate words and semantically related words, reflected in the reduced N400s, could occur primarily because these words are more plausible sentence continuations. Our data are compatible with this account, because the pattern of N400 amplitude per condition (inanimate-unrelated > inanimate-related > animate-unrelated > animate-related > plausible control) follows the pattern of the plausibility judgments (see Tables 2 and 3), with larger N400s found for increasingly implausible conditions (albeit only when participants performed the plausibility judgment task in Experiment 1). However, the observed relationship between N400 activity and plausibility does not appear to be a linear function, because big differences in plausibility between the control words and the animate-related words elicited only small N400 differences, while much smaller differences in plausibility (e.g., between the inanimate-unrelated words and the animate-related words) elicited larger N400 differences.

It is of course possible that the semantic prediction account and the integration account are both partly right (for discussion, see Van Berkum, 2009; Baggio and Hagoort, 2011; Lau et al., 2016; Nieuwland et al., 2020b). There are various demonstrations of semantic prediction that are hard to explain in terms of integration (e.g., Federmeier and Kutas, 1999; Ito et al., 2016, but see also Pickering and Gambi, 2018), and demonstrations of integration that are not straightforwardly explained in terms of prediction alone (e.g., Fleur et al., 2020; Nieuwland et al., 2020b; Wang et al., 2011). But prediction and integration are not mutually exclusive, and the effects of semantic prediction and those of integration may both impact N400 activity in a cascading manner (e.g., Nieuwland et al., 2020b). Crucially, regardless of whether our results reflect effects of prediction and/or integration, we did not obtain evidence that animacy information is used as a filter or impacts processing *before* other real-world knowledge does.

Animacy did generate larger effects than other factors: animacy violations had a larger effect on N400s than real-world knowledge violations and animacy had a larger effect on N400s than semantic relatedness. Animacy violations elicited strong effects in both experiments, whereas real-world knowledge violations elicited a relatively small effect in Experiment 1 and no clear effect in Experiment 2. At this point we cannot be certain why animacy had the strongest effect, and we note that other studies have not found such effects (Szewczyk and Schriefers, 2011; Zhang et al., 2012). It could matter that animacy violations in our study were also more implausible violations than real-world event-knowledge violations, and that the impact of animacy on plausibility was greater than that of relatedness. In this regard, animacy violations could be considered 'special' compared to other real-world knowledge violations (for discussion, see also Hung and Schumacher, 2014; Szewczyk and Schriefers, 2011). However, it is also possible that our participants simply paid extra attention to animacy, it being a salient, binary feature that occurred frequently in the experiment. The crucial point, however, is that a larger effect size in and of itself is not evidence for priority of animacy information.

Finally, we provide a note on the specific N400 effect of animate-related words compared to control words. This effect was smaller compared to effects of other violations but nevertheless statistically significant. We can only speculate why PK12 observed little evidence for such an effect. Many studies have reported no N400 effect for related-implausible or related-anomalous words (Nieuwland and Van Berkum, 2005; Van Herten, Chwilla and Kolk, 2006; for reviews, see Bornkessel-Schlesewsky and Schlesewsky, 2008; Brouwer et al., 2012; Kuperberg, 2007), but in all these cases these words elicited a P600 effect instead (the semantic P600). PK12 did not observe differential ERP activity between control words and animate-related words until the sentence-final word, and this appears to be a rather unique finding within the broader literature perhaps that their participants simply did not consistently notice anything problematic about the animate-related words until later in the sentence.[7] This may also have been the case in our Experiment 2, perhaps because participants paid less attention to sentence meaning when they did not have to judge the plausibility of each sentence.

We should also highlight that participants in PK12 always saw two conditions of each trial but never saw the control sentence first. This could have caused them to anticipate an animate and related plausible word when reading the same sentence context for the second time. This may have made it more likely that they temporarily mistook the animate-related word for a plausible word. For understanding the PK12 results, this potential confound inherent to the experimental design is relevant because the reported interaction between animacy and relatedness was for a large part driven by the reduced N400 to the animate-related condition. In our experiment, participants did not see two conditions of each item, and therefore could not have formed such a strategy based on the repetition of a sentence context.

4.2. P600 results

The third relevant pattern in our results was that, in Experiment 1, animacy violations did not elicit a P600 effect, but instead we observed enhanced anterior positivities for related compared to unrelated words and for animate compared to inanimate words. These results therefore also do not replicate the PK12 findings.

Based on their findings and previous reports (e.g., Paczynski and Kuperberg, 2011), PK12 argued that their animacy P600 effect constituted a 'semantic P600' effect (e.g., Kim and Osterhout, 2005), thought to reflect continued analysis or reanalysis that is triggered by a conflict between an expected representation and the detection of an impossible proposition. However, they could not rule out the possibility that the observed effects were due to the plausibility judgment task. While some studies found semantic P600 effects in absence of such a task (Van De Meerendonk et al., 2010; Nieuwland and Van Berkum, 2005), other studies found that selection restriction anomalies involving animacy only elicited P600 effects when participants performed a plausibility judgment task, but not when they only passively read the sentences (Kolk et al., 2003; Schacht et al., 2014).

Our own results suggest that the task played a role in the observed patterns. The P600 effects in Experiment 1 roughly patterned with task difficulty. The animate-related condition elicited the largest P600 and

[7] Like PK12, we analyzed the ERP responses to sentence-final words; those results are available online (https://osf.io/se3pc). PK12 reported that all violations elicited an enhanced N400 to the control condition and with no differences between the violation conditions. From this they argued that all violation conditions ultimately evoked similar downstream processing consequences. Our own results for Experiment 1 showed extended N400-like activity for all violation conditions compared to control (all $ps < .001$). In Experiment 2, only inanimate-unrelated words elicited significantly extended N400-like activity compared to control words ($p = .029$). Sentence-final figures provided in our online supplementary materials also show the different observed patterns with and without task in our two experiments.

was also the most difficult, as evident from the response accuracy and response time measurements. This condition may be more difficult precisely because these words are both animate and related to the context. Moreover, in Experiment 2, where there was no plausibility judgment task, we did not observe any P600 patterns, and, if anything, animacy violations elicited extended N400 activity. In this experiment, the ERP waveforms across all conditions were very different, namely less positive going, from those obtained in Experiment 1.

We think that the P600 effects in our Experiment 1, and possibly those in PK12, reflect the concomitant ERP effects of the task (i.e., decision-based effects). We note that sensitivity to task-demands of the P600 effects does not necessarily contradict the interpretation of PK12 that the P600 activity reflected detection of propositional impossibility. In their study, participants may have considered inanimate words as especially relevant to the task, despite the balanced design with filler sentences so that inanimate nouns were encountered equally often in plausible and implausible sentences. However, we think that these differences between studies and experiments serve as a warning about always or only using judgment tasks in such studies. Judgment tasks indeed sometimes boost engagement with the linguistic materials, yet they can change the comprehension processes of interest, and they can elicit ERP components that would otherwise not be elicited, potentially masking N400 activity, because of spatio-temporal component overlap.

Additional analyses comparing the patterns observed in Experiment 1 and 2 by the addition of the between-subject 2-level factor experiment type (task, no task) can be found in our supplementary materials (https://osf.io/se3pc). Here we provide a brief discussion of the most salient patterns. The N400 effects of animacy did not differ when participants performed an acceptability judgment task from when they did not. However, the task did affect semantic relatedness, because we observed an effect of semantic relatedness in Experiment 1 but not in Experiment 2. One possible conclusion is that participants are sensitive to semantic relatedness only when performing an acceptability judgment task. However, the effect of relatedness in Experiment 1 was not as strong as the effect of animacy, which would make it harder to detect without the task, because the task allowed us to select only the trials where participants' responses matched to the expected plausibility ratings of the experimental conditions. We know from various other studies that semantically related words do lead to reduced N400 effects in the absence of such a task, although these studies typically present constrained sentences that led to an expectation of a certain word (e.g., Metusalem et al., 2012; Nieuwland, 2016; Ito et al., 2016). However, it is also entirely possible that the absence of a semantic relatedness effect in this experiment was not detected because of low statistical power, given the small sample size in Experiment 2.

Another salient difference between the experiments was in the post-N400 time window. While all conditions in Experiment 1 elicited strongly positive-going ERPs in this window, the N400 effects in Experiment 2 extended beyond the typical N400 time window, leading to a sustained negativity and absence of clear post-N400 positive deflections. In Experiment 2, we thus did not observe the anterior positivities for related compared to unrelated conditions and for animate compared to inanimate conditions seen in Experiment 1. Based on such results, we conclude that the post-N400 activity in Experiment 1 was likely a result of the acceptability task (see also Kolk et al., 2003).

4.3. No evidence for priority of animacy during sentence comprehension

Our results only partially replicate the findings observed by PK12. In the N400 window, we found strong effects of animacy in Experiments 1 and 2, and a relatively small effect of relatedness in Experiment 1. But we did not observe the interaction pattern between animacy and relatedness found by PK12. Although the lack of an interaction effect is a null result, we note that this pattern arose from observing effects where PK12 did not: a small relatedness effect for inanimate words, and an animacy effect for unrelated words. Moreover, we also observed an N400 difference between animate-related words and control words where PK12 did not.

At the P600 window, in Experiment 1 only we found enhanced anterior positivities for animates compared to inanimates and for related compared to unrelated words, but not a posterior P600 animacy effect. We used the exact same design as PK12 and used their materials, but we adapted and expanded the original set of materials, and our participants never saw two conditions of the same item, unlike in PK12. Because we tested twice the number of participants as PK12, the number of observations for analysis was roughly the same (but our results are more likely to generalize to new participant observations because we modeled more subject-variance). PK12 used 120 experimental sentence contexts twice in two different conditions, rendering the experiment very long, and making participants tired, leading to potential repetition effects, and perhaps inducing experiment-specific strategies. We wanted to reduce the length of the experiment, and avoid within-participant repetition of sentence contexts to minimize potential processing strategies. Whether and how these differences explain the discrepancy between the observed results remains an open question. Perhaps more importantly, it remains to be seen whether the results of PK12 and/or those reported here are replicable (to encourage replication, we made all our materials and presentation scripts available online on https://osf.io/se3pc), but we note that we observed the same pattern of results in our first twenty participants and the twenty additional participants we tested (for information, see footnote 2).

In conclusion, in contrast to PK12, our results support an incremental, constraint-based account of semantic processing (MacDonald et al., 1994; Marslen-Wilson et al., 1988). At least at the point where the meaning of a word is retrieved, animacy and real-world event-knowledge appear to be used concurrently, in a similarly incremental fashion.

Author note

All data, stimulus materials and data analysis files were available during peer review and remain publicly available on the Open Science Framework (https://osf.io/se3pc), in accordance with the Peer Review Openness Initiative (http://opennessinitiative.org; Morey et al., 2016).

Credit author statement

MVM: Writing, review & editing, methodology, investigation, formal analysis, project administration. MJP: Writing, review & editing, supervision. MSN: Writing, review & editing, methodology, formal analysis, supervision, project administration.

Acknowledgements

This research was supported by PPLS Research Support Grants. The first author would also like to thank CONACYT, Mexico and The University of Edinburgh for supporting her doctoral research. The first author was also supported by a Knut and Alice Wallenberg Foundation grant (No. KAW 2014.0205). We also thank Jakub M. Szewczyk for helpful comments on our manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2020.107724.

Appendix I

Reanalyzed data results

Upon reviewer request, we re-analyzed our data using a 0.1–20 Hz filter, corrected channels only using topographic interpolation, and using a more focused statistical approach. Our reanalysis therefore focuses on the medial ROI. Furthermore, because both PK and our own results suggest that the observed N400 effects are broadly distributed within that ROI, we excluded distributional factors from the N400 analysis and examined activity averaged over all channels in the medial ROI. For the P600 time window, because both PK and our own results suggest that the observed effects depend on anteriority, we included the anteriority factor for these analyses.

These re-analyses yielded highly similar results as the original analyses described in the main text. The only difference was that in these new analyses, but not the original analyses, the enhanced N400 for animate-related words compared to control words in Experiment 2 was statistically significant, as we had also observed in Experiment 1. The average numbers of computed trials per condition used in the new analyses were also similar to those in the original analyses.

Experiment 1

N400 Results

Pairwise (control vs. violation type) analyses. A series of repeated measures General Linear Models (GLMs – ANOVAs) comparing the control condition with each of the violation types showed that control words elicited significantly smaller (more positive) N400s ($M = 3.71$ μV, $SD = 3.46$) than animate-related ($M = 2.39$ μV, $SD = 3.22$, $M_{diff} = 1.32$ μV, $\eta_p^2 = 0.126$), animate-unrelated ($M = 1.31$ μV, $SD = 3.13$, $M_{diff} = 2.40$ μV, $\eta_p^2 = 0.473$), inanimate-related ($M = 0.21$ μV, $SD = 3.21$, $M_{diff} = 3.50$ μV, $\eta_p^2 = 0.604$), and inanimate-unrelated words ($M = -0.48$ μV, $SD = 2.98$, $M_{diff} = 4.20$ μV, $\eta_p^2 = 0.674$) (Table A1).

Table A1
F-value results from the pairwise ANOVAs contrasting each type of violation with the control condition at different time windows.

	N400 (300–500 ms)	P600 (700–900 ms)	
	Sent	Sent	Sent x AP
(2) Animate-Related	5.6*	7.6**	3.1^
(3) Animate-Unrelated	34.9***	.78	1.9
(4) Inanimate-Related	59.6***	.47	2.6
(5) Inanimate-Unrelated	80.7***	1.3	50.0***

Notes. DF [1,39]. Sent: Sentence Type. AP: Anterior-Posterior distribution.
^$p < .1$
*$p < .05$
**$p < .01$
***$p < .001$

Animacy by relatedness interaction analyses. A repeated measures ANOVA with factors animacy (animate, inanimate) and relatedness (related, unrelated) showed main effects of animacy and relatedness (Table A2) such that animate words elicited smaller N400s ($M = 1.85$ μV, $SD = 2.94$) than inanimate words ($M = -0.14$ μV, $SD = 2.89$, $M_{diff} = 1.99$ μV, $\eta_p^2 = 0.517$) and likewise, related words elicited smaller N400s ($M = 1.30$ μV, $SD = 3.00$) than unrelated words ($M = 0.41$ μV, $SD = 2.78$, $M_{diff} = 0.89$ μV, $\eta_p^2 = 0.199$). The interaction between animacy and relatedness was not significant ($F < 1$) (Table A2).

Table A2
F-value results from the 2 × 2 ANOVAs at the medial column

	Main effect of Ani	Main effect of Rela	Ani x Rela	Ani xAP	Rela x AP	Ani x Rela x AP
N400 (300–500 ms)	41.8***	9.7**	.66	–	–	–
P600 (700–900 ms)	2.5	2.2	3.1^	32.8***	33.6***	3.5^

Notes. DF [1,39]. Ani: Animacy. Rela: Relatedness. AP: Anterior-Posterior distribution.
^$p < .1$
*$p < .05$
**$p < .01$
***$p < .001$

P600 Results

Pairwise (control vs. violation type) analyses. A series of repeated measures General Linear Models (GLMs – ANOVAs) comparing the control condition with each of the violation type and including the anteriority factor (anterior, posterior) showed that animate-related words elicited more positive ERPs ($M = 6.56$ μV, $SD = 3.62$) than the control condition ($M = 4.96$ μV, $SD = 3.15$) as indicated by a main effect of sentence type ($M_{diff} = -1.60$ μV, $\eta_p^2 = 0.163$). The interaction between sentence and anteriority for this comparison was only marginal (Table A1).

For the comparison between the inanimate-unrelated and the control conditions, the main effect of sentence type was not significant, however, the sentence by anteriority interaction was significant (Table A1). Follow up of this interaction showed that inanimate-unrelated words elicited more

positive ERPs than the control condition at the posterior region (control posterior: $M = 4.87$ μV, $SD = 3.44$, inanimate-unrelated posterior: $M = 6.71$ μV, $SD = 3.17$, $M_{diff} = -1.83$ μV, $F_{1,39} = 12.6$, $p = .001$, $\eta_p^2 = 0.244$), while the difference between the two sentence types did not reach statistical significance at the anterior region (control anterior: $M = 5.04$ μV, $SD = 3.23$, inanimate-unrelated anterior: $M = 4.23$ μV, $SD = 2.81$, $M_{diff} = 0.80$ μV, $F_{1,39} = 2.9$, $p = .096$, $\eta_p^2 = 0.069$). Finally, neither the main effect of sentence type nor the interaction between sentence type and anteriority were significant for the animate-unrelated and inanimate-related comparisons to the control condition (Table A1).

Animacy by relatedness interaction analyses. As table A2 shows, we did not find main effects of animacy or relatedness ($ps \geq .122$). However, the 2-way animacy by anteriority interaction revealed more positive ERPs for animates than inanimates at anterior (Animates $M = 6.12$ μV, $SD = 3.75$; Inanimates $M = 4.67$ μV, $SD = 2.76$, $M_{diff\ =} 1.45$ μV, $F_{1,39} = 10.2$, $p = .003$, $\eta_p^2 = 0.208$) but not at posterior regions (Animates $M = 5.88$ μV, $SD = 3.47$; Inanimates $M = 6.09$ μV, $SD = 3.12$, $M_{diff\ =} -0.21$ μV, $F < 1$). Similarly, the significant relatedness by anteriority interaction showed more positive ERPs for related than unrelated words at anterior (Related $M = 6.04$ μV, $SD = 2.97$; Unrelated $M = 4.75$ μV, SD = 3.31, $M_{diff} = 1.29$ μV, $F_{1,39} = 15.0$, $p < .001$, $\eta_p^2 = 0.278$) but not at posterior regions (Related $M = 5.80$ μV, $SD = 3.39$; Unrelated $M = 6.16$ μV, $SD = 3.14$, $M_{diff} = -0.36$ μV, $F_{1,39} = 1.1$, $p = .307$, $\eta_p^2 = 0.027$). The 2-way interaction between animacy and relatedness and the 3-way interaction between animacy, relatedness and anteriority were only marginal (Table A2).

Summary of Experiment 1. At the N400 time-window, control words elicited significantly smaller N400s than all violation sentences. In the animacy by relatedness analysis, animate words elicited smaller N400s than inanimate words, and related words elicited smaller N400s than unrelated words. The interaction between animacy and relatedness was not significant.

Results from the P600 time-window showed more positive ERPs overall for animate-related than control words and more positive ERPs for inanimate-unrelated than control words at the posterior region. The animacy by relatedness interaction analysis showed more positive ERPs for animate than inanimate words at anterior regions and more positive ERPs for related than unrelated words also at anterior regions.

Experiment 2 No Task

N400 Results

Pairwise (control vs. violation type) analyses. A series of repeated measures General Linear Models (GLMs – ANOVAs) comparing the control condition with each of the violation types showed that control words elicited significantly smaller (more positive) N400s ($M = 1.02$ μV, $SD = 2.60$) than animate-related ($M = -0.14$ μV, $SD = 2.51$, $M_{diff} = 1.16$ μV, $\eta_p^2 = 0.230$), inanimate-related ($M = -1.04$ μV, $SD = 2.22$, $M_{diff} = 2.06$ μV, $\eta_p^2 = 0.375$), and inanimate-unrelated words ($M = -1.40$ μV, $SD = 4.07$, $M_{diff} = 2.42$ μV, $\eta_p^2 = 0.359$). The difference between control and animate-unrelated words was only marginally significant (animate-unrelated $M = 0.10$ μV, $SD = 2.47$, $M_{diff} = 0.92$ μV, $\eta_p^2 = 0.195$) (Table A3).

Table A3
F-value results from the pairwise ANOVAs contrasting each type of violation with the control condition at different time windows.

	N400 (300–500 ms)	P600 (700–900 ms)	
	Sent	Sent	Sent x AP
(2) Animate-Related	4.8*	1.8	1.4
(3) Animate-Unrelated	3.9ˆ	.05	.18
(4) Inanimate-Related	9.6**	.04	.38
(5) Inanimate-Unrelated	9.0**	.15	1.9

Notes. DF [1,16]. Sent: Sentence Type. AP: Anterior-Posterior distribution.
ˆ$p < .1$
*$p < .05$
**$p < .01$
*** $p < .001$

Animacy by relatedness interaction analyses. The significant main effect of animacy showed that animate words elicited smaller N400s ($M = -0.02$ μV, $SD = 2.31$) than inanimate words ($M = -1.22$ μV, $SD = 2.69$, $M_{diff} = 1.20$ μV, $\eta_p^2 = 0.334$). Neither the main effect of relatedness, nor the interaction between animacy and relatedness were significant ($Fs < 1$) (Table A4).

Table A4
F-value results from the 2 × 2 ANOVAs at the medial column

	Main effect of Ani	Main effect of Rela	Ani x Rela	Ani x AP	Rela x AP	Ani x Rela x AP
N400 (300–500 ms)	8.0*	.02	.31	–	–	–
P600 (700–900 ms)	1.5	.63	.14	2.3	2.0	.64

Notes. DF [1,16]. Ani: Animacy. Rela: Relatedness. AP: Anterior-Posterior distribution.
ˆ$p < .1$
*$p < .05$
**$p < .01$
***$p < .001$

P600 Results

Pairwise (control vs. violation type) analyses. None of the comparisons between control words violation words was statistically significant (*ps* ≥ .201) and none of the interactions between sentence type and anteriority were significant either (*ps* ≥ .191) (Table A3).

Animacy by relatedness interaction analyses. No significant main effects or interactions were found in this analysis (*ps* ≥ .150) (Table A4).

Summary of Experiment 2

At the N400 time-window, in the pairwise comparisons, control words elicited smaller N400s than animate-related, inanimate-related and inanimate-unrelated words. In the animacy by relatedness interaction analyses, animate words elicited smaller N400s than inanimate words.

Results from the P600 time-window did not yield any statistically significant differences between the control and the violation conditions. No significant main effects or interactions were found in the animacy by relatedness analysis at this time-window.

References

Altmann, G., Mirković, J., 2009. Incrementality and prediction in human sentence processing. Cognit. Sci. 33 (4), 583–609.

Baggio, G., Hagoort, P., 2011. The balance between memory and unification in semantics: a dynamic account of the N400. Lang. Cognit. Process. 26 (9), 1338–1367.

Bornkessel-Schlesewsky, I., Schlesewsky, M., 2008. An alternative perspective on "semantic P600" effects in language comprehension. Brain Res. Rev. 59 (1), 55–73. https://doi.org/10.1016/j.brainresrev.2008.05.003.

Branigan, H., Pickering, M., Tanaka, M., 2008. Contributions of animacy to grammatical function assignment and word order during production. Lingua 118 (2), 172–189.

Brouwer, H., Fitz, H., Hoeks, J., 2012. Getting real about Semantic Illusions: rethinking the functional role of the P600 in language comprehension. Brain Res. 1446, 127–143. https://doi.org/10.1016/j.brainres.2012.01.055.

Brown, C., Hagoort, P., 1993. The processing nature of the N400: evidence from masked priming. J. Cognit. Neurosci. 5 (1), 34–44.

Chwilla, D., Brown, C., Hagoort, P., 1995. The N400 as a function of the level of processing. Psychophysiology 32 (3), 274–285.

Clark, H., 1965. Some structural properties of simple active and passive sentences. J. Verb. Learn. Verb. Behav. 4 (5), 365–370.

De Swart, P., Van Bergen, G., 2019. How animacy and verbal information influence V2 sentence processing: evidence from eye movements. Open Ling. 5 (1), 630–649.

Ehrlich, S.F., Rayner, K., 1981. Contextual effects on word perception and eye movements during reading. J. Verb. Learn. Verb. Behav. 20 (6), 641–655.

Federmeier, K., Kutas, M., 1999. A rose by any other name: long-term memory structure and sentence processing. J. Mem. Lang. 41 (4), 469–495.

Ferretti, T.R., Kutas, M., McRae, K., 2007. Verb aspect and the activation of event knowledge. J. Exp. Psychol. Learn. Mem. Cognit. 33 (1), 182.

Filik, R., Leuthold, H., 2008. Processing local pragmatic anomalies in fictional contexts: evidence from the N400. Psychophysiology 45 (4), 554–558.

Fleur, D.S., Flecken, M., Rommers, J., Nieuwland, M.S., 2020. Definitely saw it coming? The dual nature of the pre-nominal prediction effect. Cognition 204, 104335.

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

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

Greenhouse, S., Geisser, S., 1959. On methods in the analysis of profile data. Psychometrika 24 (2), 95–112.

Hung, Y.C., Schumacher, P.B., 2014. Animacy matters: ERP evidence for the multi-dimensionality of topic-worthiness in Chinese. Brain Res. 1555, 36–47.

Ito, A., Corley, M., Pickering, M.J., Martin, A.E., Nieuwland, M.S., 2016. Predicting form and meaning: evidence from brain potentials. J. Mem. Lang. 86, 157–171. https://doi.org/10.1016/j.jml.2015.10.007.

Ito, A., Martin, A.E., Nieuwland, M.S., 2017. On predicting form and meaning in a second language. J. Exp. Psychol. Learn. Mem. Cognit. 43 (4), 635–652. https://doi.org/10.1037/xlm0000315.

JASP Team, 2017. JASP ([Computer software]), Version 0.8.1.

Jeffreys, H., 1961. Theory of Probability, third ed. Oxford University Press.

Kim, A., Osterhout, L., 2005. The independence of combinatory semantic processing: evidence from event-related potentials. J. Mem. Lang. 52 (2), 205–225.

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

Kuperberg, G., 2007. Neural mechanisms of language comprehension: challenges to syntax. Brain Res. 1146, 23–49.

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

Landauer, T.K., 1998. Learning and representing verbal meaning: the latent semantic analysis theory. Curr. Dir. Psychol. Sci. 7 (5), 161–164.

Landauer, T., Foltz, P., Laham, D., 1998. An introduction to latent semantic analysis. Discourse Process 25 (2–3), 259–284.

Lau, E.F., Namyst, A., Fogel, A., Delgado, T., 2016. A direct comparison of N400 effects of predictability and incongruity in adjective-noun combination. Collabra 2 (1).

MacDonald, M.C., Seidenberg, M.S., 2006. Constraint satisfaction accounts of lexical and sentence comprehension. Handbook of Psycholinguistics. Academic Press, pp. 581–611.

MacDonald, M.C., Pearlmutter, N.J., Seidenberg, M.S., 1994. The lexical nature of syntactic ambiguity resolution. Psychol. Rev. 101 (4), 676.

Marslen-Wilson, W., 1973. Linguistic structure and speech shadowing at very short latencies. Nature 244 (5417), 522–523. https://doi.org/10.1038/244522a0.

Marslen-Wilson, W., Tyler, L., 1980. The temporal structure of spoken language understanding. Cognition 8 (1), 1–71.

Marslen-Wilson, W., Brown, C.M., Tyler, L.K., 1988. Lexical representations in spoken language comprehension. Lang. Cognit. Process. 3 (1), 1–16. https://doi.org/10.1080/01690968808402079.

Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., Sommer, W., 2006. Semantics prevalence over syntax during sentence processing: a brain potential study of noun–adjective agreement in Spanish. Brain Res. 1093 (1), 178–189.

Matsuki, K., Chow, T., Hare, M., Elman, J., 2011. Event-based plausibility immediately influences on-line language comprehension. J. Exp. Psychol. Learn. Mem. Cognit. 37 (4), 913.

McRae, K., Matsuki, K., 2009. People use their knowledge of common events to understand language, and do so as quickly as possible. Language and Linguistics Compass 3 (6), 1417–1429.

Metusalem, R., Kutas, M., Urbach, T., Hare, M., 2012. Generalized event knowledge activation during online sentence comprehension. J. Mem. Lang. 66 (4), 545–567.

Milburn, E., Warren, T., Dickey, M.W., 2016. World knowledge affects prediction as quickly as selectional restrictions: evidence from the visual world paradigm. Language, Cognition and Neuroscience 31 (4), 536–548.

Morey, R.D., Rouder, J.N., 2015. BayesFactor ([Computer software]), Version 0.9.10-2.

Morey, R.D., Chambers, C.D., Etchells, P.J., Harris, C.R., Hoekstra, R., Lakens, D., et al., 2016. The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review. Royal Society Open Science 3 (1), 150547.

Nieuwland, M.S., 2014. "Who's he?" Event-related brain potentials and unbound pronouns. J. Mem. Lang. 76, 1–28. https://doi.org/10.1016/j.jml.2014.06.002.

Nieuwland, M., 2015. The truth before and after: brain potentials reveal automatic activation of event knowledge during sentence comprehension. J. Cognit. Neurosci. 27 (11), 2215–2228.

Nieuwland, M.S., 2016. Quantification, prediction, and the online impact of sentence truth-value: evidence from event-related potentials. J. Exp. Psychol. Learn. Mem. Cognit. 42 (2), 316–334. https://doi.org/10.1037/xlm0000173.

Nieuwland, M.S., 2019. Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. Neurosci. Biobehav. Rev. 96, 367–400.

Nieuwland, M.S., Van Berkum, J.J.A., 2005. Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. Cognit. Brain Res. 24 (3), 691–701.

Nieuwland, M.S., Van Berkum, J.J.A., 2006. When peanuts fall in love: N400 evidence for the power of discourse. J. Cognit. Neurosci. 18 (7), 1098–1111.

Nieuwland, M.S., Martin, A.E., Carreiras, M., 2013. Event-related brain potential evidence for animacy processing asymmetries during sentence comprehension. Brain Lang. 126 (2), 151–158.

Nieuwland, M.S., Arkhipova, Y., Rodríguez-Gómez, P., 2020a. Anticipating words during spoken discourse comprehension: A large-scale, pre-registered replication study using brain potentials. Cortex 133, 1–36.

Nieuwland, M.S., Barr, D.J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D.I., Ferguson, H.J., Fu, X., Heyselaar, E., Huettig, F., Husband, E.M., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., Rueschemeyer, S.-A., Segaert, K., Tuomainen, J., Von Grebmer Zu Wolfsthurn, S., 2020b. Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 375, 20180522 https://doi.org/10.1098/rstb.2018.0522.

Osterhout, L., 1999. A superficial resemblance does not necessarily mean you are part of the family: counterarguments to coulson, king and Kutas (1998) in the P600/SPS-P300 debate. Lang. Cognit. Process. 14 (1), 1–14. https://doi.org/10.1080/016909699386356.

Osterhout, L., Mobley, L., 1995. Event-related brain potentials elicited by failure to agree. J. Mem. Lang. 34 (6), 739–773.

Otten, M., Berkum, J. Van, 2008. Discourse-based word anticipation during language processing: prediction or priming? Discourse Process 45 (6), 464–496.

Paczynski, M., Kuperberg, G., 2011. Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb-argument processing. Lang. Cognit. Process. 26 (9), 1402–1456.

Paczynski, M., Kuperberg, G.R., 2012. Multiple influences of semantic memory on sentence processing: distinct effects of semantic relatedness on violations of real-

world event/state knowledge and animacy selection restrictions. J. Mem. Lang. 67 (4), 426–448.
Pickering, M.J., Gambi, C., 2018. Predicting while comprehending language: a theory and review. Psychol. Bull. 144 (10), 1002.
Pickering, M., Traxler, M., 1998. Plausibility and recovery from garden paths: an eye-tracking study. J. Exp. Psychol. Learn. Mem. Cognit. 24 (4), 940.
Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., Schlesewsky, M., 2007. To predict or not to predict: influences of task and strategy on the processing of semantic relations. J. Cognit. Neurosci. 19 (8), 1259–1274.
Rouder, J.N., Morey, R.D., Speckman, P.L., Province, J.M., 2012. Default Bayes factors for ANOVA designs. J. Math. Psychol. 56 (5), 356–374.
Sassenhagen, J., Bornkessel-Schlesewsky, I., 2015. The P600 as a correlate of ventral attention network reorientation. Cortex 66, A3–A20.
Sassenhagen, J., Schlesewsky, M., 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 Lang. 137, 29–39.
Schacht, A., Sommer, W., Shmuilovich, O., Martíenz, P.C., Martín-Loeches, M., 2014. Differential task effects on N400 and P600 elicited by semantic and syntactic violations. PloS One 9 (3), e91226.
Schumacher, P.B., 2012. Context in neurolinguistics: time-course data from electrophysiology. In: Finkbeiner, R., Meibauer, J., Schumacher, P.B. (Eds.), What Is a Context? Linguistic Approaches and Challenges. John Benjamins, Amsterdam, pp. 33–53.
Schwanenflugel, P.J., Shoben, E.J., 1985. The influence of sentence constraint on the scope of facilitation for upcoming words. J. Mem. Lang. 24 (2), 232–252.
Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., 2007. The time course of plausibility effects on eye movements in reading: evidence from noun-noun compounds. J. Exp. Psychol. Learn. Mem. Cognit. 33 (6), 1162.
Szewczyk, J.M., Schriefers, H., 2011. Is animacy special?: ERP correlates of semantic violations and animacy violations in sentence processing. Brain Res. 1368, 208–221.
Szewczyk, J.M., Schriefers, H., 2013. Prediction in language comprehension beyond specific words: an ERP study on sentence comprehension in Polish. J. Mem. Lang. 68 (4), 297–314.
Trueswell, J.C., Tanenhaus, M.K., Garnsey, S.M., 1994. Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. J. Mem. Lang. 33 (3), 285.
Van Berkum, J.J., 2009. The neuropragmatics of 'simple' utterance comprehension: an ERP review. Semantics and Pragmatics: from Experiment to Theory. Palgrave Macmillan, pp. 276–316.
Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. J. Exp. Psychol. Learn. Mem. Cognit. 31 (3), 443.
Van De Meerendonk, N., Kolk, H.H., Vissers, C.T.W., Chwilla, D.J., 2010. Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. J. Cognit. Neurosci. 22 (1), 67–82.
Van Herten, M., Chwilla, D.J., Kolk, H.H., 2006. When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. J. Cognit. Neurosci. 18 (7), 1181–1197.
Van Petten, C., Luka, B., 2012. Prediction during language comprehension: benefits, costs, and ERP components. Int. J. Psychophysiol. 83 (2), 176–190.
Wang, L., Bastiaansen, M., Yang, Y., Hagoort, P., 2011. The influence of information structure on the depth of semantic processing: how focus and pitch accent determine the size of the N400 effect. Neuropsychologia 49 (5), 813–820. https://doi.org/10.1016/j.neuropsychologia.2010.12.035.
Warren, T., McConnell, K., 2007. Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. Psychon. Bull. Rev. 14 (4), 770–775.
Weckerly, J., Kutas, M., 1999. An electrophysiological analysis of animacy effects in the processing of object relative sentences. Psychophysiology 36 (5), 559–570.
West, W., Holcomb, P., 2000. Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. J. Cognit. Neurosci. 12 (6), 1024–1037.
Wicha, N.Y., Moreno, E.M., Kutas, M., 2004. Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. J. Cognit. Neurosci. 16 (7), 1272–1288.
Zhang, Y., Zhang, J., Min, B., 2012. Neural dynamics of animacy processing in language comprehension: ERP evidence from the interpretation of classifier–noun combinations. Brain Lang. 120 (3), 321–331.