



## “If a lion could speak . . .”: Online sensitivity to propositional truth-value of unrealistic counterfactual sentences

Mante S. Nieuwland

Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK  
 Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69-2, 20009 Donostia-San Sebastian, Spain

### ARTICLE INFO

#### Article history:

Received 2 April 2012  
 revision received 27 August 2012  
 Available online 26 September 2012

#### Keywords:

Language comprehension  
 Counterfactual conditionals  
 Propositional truth-value  
 Real-world knowledge  
 Event-related potentials  
 N400

### ABSTRACT

People can establish whether a sentence is hypothetically true even if what it describes can never be literally true given the laws of the natural world. Two event-related potential (ERP) experiments examined electrophysiological responses to sentences about unrealistic counterfactual worlds that require people to construct novel conceptual combinations and infer their consequences as the sentence unfolds in time (e.g., “If dogs had gills. . .”). Experiment 1 established that without this premise, described consequences (e.g., “Dobermans would breathe under water . . .”) elicited larger N400 responses than real-world true sentences. Incorporation of the counterfactual premise in Experiment 2 generated similar N400 effects of propositional truth-value in counterfactual and real-world sentences, suggesting that the counterfactual context eliminated the interpretive problems posed by locally anomalous sentences. This result did not depend on cloze probability of the sentences. In contrast to earlier findings regarding online comprehension of logical operators and counterfactuals, these results show that ongoing processing can be directly impacted by propositional truth-value, even that of unrealistic counterfactuals.

© 2012 Elsevier Inc. All rights reserved.

### Introduction

Wittgenstein (1958) famously argued that because language use is fundamentally grounded in common patterns of human behavior and cognition “If a lion could speak, we could not understand him”. His thought experiment illustrates the capacity to reason counterfactually about events that can never realistically take place: we can consider whether his abstruse conclusion is hypothetically true despite knowing that what it describes can never be literally true given the biological or physical laws of the world. But does propositional truth-value impact our comprehension of sentences, including ‘unrealistic’ counterfactuals, as they unfold in time? This paper reports two event-related potential (ERP) studies that investigated whether the language processing system is immediately sensitive to the truth-value of unrealistic counterfactual conditionals such

as “If dogs had gills, Dobermans would breathe under water”.

Balancing our knowledge of what is true and possible in the world with what we can experience in thought and express in language may be particularly taxing when we process language with unrealistic content (as compared to, let’s say, fictional yet realistic content). Understanding an unrealistic counterfactual sentence requires the creation of an unusual conceptual combination (‘dogs’ and ‘gills’) with potential consequences that go beyond what is literally stated (e.g., the hobby of keeping dogs as pets would drastically change if they had gills instead of lungs). Whereas the sentence “Dobermans would breathe under water” by itself semantically anomalous, it may describe a hypothetically true consequence of the counterfactual premise. However, is our language comprehension system immediately sensitive to its truth-value, as compared to that of a more realistic sentence that draws upon our pre-existing real-world knowledge (e.g., “Because fish have gills, tuna breathe under water”)? And does a plausible and

E-mail address: [m.nieuwland@ed.ac.uk](mailto:m.nieuwland@ed.ac.uk)

considered-to-be-true relationship between counterfactual consequence and premise mitigate processing costs associated with a sentence that is false with respect to knowledge of the real-world?

This latter question resonates with classic empirical questions regarding the interaction between local and global levels of language comprehension (e.g., word- or phrase-level versus discourse-level). Experimental results suggest that supportive discourse context can mitigate processing costs associated with phrases or sentences that describe unexpected events (e.g., “the electrician taught herself”; Cook & Myers, 2004; Duffy & Keir, 2004; Hess, Foss, & Carroll, 1995). According to the lexical reinterpretation model (e.g., Hess et al., 1995), discourse context modifies the representation of a concept to form the basis for further interpretation (see also Zwaan & Radvansky, 1998). Such results testify to an interactive view of language processing in which lexical, structural, and contextual information interact efficiently as a message unfolds (e.g., Marslen-Wilson, 1975), and are consistent with contextually-driven resolution (or preclusion) of syntactic ambiguity (e.g., Altmann & Steedman, 1988) and lexical ambiguity (e.g., Rayner, Cook, Juhasz, & Frazier, 2006), and with the absence of additional processing costs for enriched expressions in supportive context (e.g., Gerrig & Bortfeld, 1999; Gibbs, 1979; Traxler, McElree, Williams, & Pickering, 2005). However, it is clear neither how much discourse context is needed to modify an initial – or ‘default’ – interpretation, nor how explicit a context must be. The current study aims to establish whether a single supportive counterfactual clause can suffice.

Whereas several studies have reported that contextual relevance outweighs real-world constraints from an early moment on (e.g., a cartoon-like story about an amorous peanut can invert the relative ease of processing “the peanut was salted” as compared to “the peanut in love”, Nieuwland & Van Berkum, 2006; see also Filik, 2008; Filik & Leuthold, 2008), some results on the processing of literally false or unrealistic sentences suggest that context does not completely overrule briefly disruptive effects of local violations (e.g., Hald, Steenbeek-Planting, & Hagoort, 2007; Warren, McConnell, & Rayner, 2008). For example, Warren et al. reported that despite a fantasy-context (e.g., Harry Potter practicing magic spells on food items), words that incurred semantic violations (e.g., ‘bread’ in “Harry used a book to teach the tough bread”) elicited longer fixations and gaze durations than in unproblematic sentences (e.g., “Harry used a microwave to heat the tough bread”). These brief disruptions correspond with predictions from the Bonding and Resolution framework (e.g., Garrod & Terras, 2000) and ‘memory-based’ language processing theories (e.g., Gerrig & O’Brien, 2005; Kintsch, 1988; Myers & O’Brien, 1998) that initial processing is dominated by local semantic relationships with effects of contextual relevance or propositional truth-value emerging later.

There is little reason to believe that the reported discrepancies stem from different techniques (for example, if eye-tracking during reading were sensitive to early processes that ERPs, in particular the N400 amplitude dependent measure, are not). Ferguson, Sanford, and Leuthold (2008) reported that early processing disruptions as in-

dexed by increased first-pass reading times but not in later measures coincided with an increased N400. This is consistent with observations that first fixation durations are correlated with N400 amplitude and that these measures are sensitive to the same lexical and sentence variables (e.g., Dambacher & Kliegl, 2007; see Dimigen, Sommer, Hohlfield, Jacobs, & Kliegl, 2011, for in-depth discussion). Hence, ERPs and eye-tracking can, in principle, be similarly sensitive to the type of processing disruptions during language comprehension under study here, despite the fact that serial visual presentation as used in ERP research on sentence reading is less natural than sentence reading eye-tracking.<sup>1</sup>

The reported discrepancies may, however, reflect differences between studies in how smoothly interpreting incoming information is related to the context, perhaps depending on how elaborate and explicit the discourse context is (see Warren et al., 2008, for discussion). Warren et al. created a variety of fantasy contexts that each made a novel unrealistic event more plausible, but realistic continuations were nevertheless on average more expected than unrealistic continuations (for related discussion, see Matsuki et al., 2011). In contrast, Nieuwland and Van Berkum (2006) created relatively uniform cartoon-like contexts that always involved an inanimate object behaving like a human being, and that each consisted of 4–5 sentences containing repeated animacy violations. This experimental manipulation strongly constrained the interpretation of the critical phrases, but it begs the questions whether the reported results were contingent upon repeated exposure of similar anomalies on the same lexical item, as could be argued from the lexical reinterpretation account (e.g., Hess et al., 1995), or whether they reflect expectations based on the discourse genre and therefore carry over to novel entities and events (e.g., Zwaan, 1994; see also Gerrig & Murphy, 1992; Van Berkum, 2010). The current studies address this by examining comprehension of single-sentence counterfactual conditionals about events that do not correspond with the natural world and that require readers to incrementally compute the unrealistic consequences of a counterfactual premise.

To understand counterfactual language requires that people balance their factual knowledge about the world with their readiness to engage in suspension of disbelief (e.g., Searle, 1975). Counterfactual thought enables humans to decouple from reality (e.g., Cosmides & Tooby, 2000), yet may be organized along the same principles as rational thought, and some aspects of reality are more readily ‘undone’ than others (see Byrne, 2007, for review). People do not usually create “miracle-world” counterfactuals but generate plausible alternatives to real-world

<sup>1</sup> Word-by-word presentation is standard in ERP research to prevent eye-movement-induced artefacts in the EEG signal (but see Dimigen et al., 2011). However, because participants cannot backtrack as people do in reading, the presentation rate is adjusted to allow participants to read comfortably. Whereas standard word-by-word presentation is slower than current estimates of natural reading pace, it elicits N400 effects that are quantitatively and qualitatively similar to those found for auditory presentation (e.g., Van Berkum, Zwitserlood, Hagoort, & Brown, 2003; Van Berkum, Hagoort, & Brown, 1999). Importantly, speeded presentation rates (e.g., 350 ms per word) are reported as uncomfortable and impair sentence comprehension (e.g., Camblin et al., 2007).

situations (e.g., McMullen & Markman, 2002). Nevertheless, the ability to generate and understand unrealistic counterfactuals testifies to the creative nature of our cognitive endowment (e.g., Fauconnier & Turner, 2002).

Recent studies have looked at counterfactual language comprehension using online measures (de Vega, Urrutia, & Rizzo, 2007; Ferguson, 2012; Ferguson & Sanford, 2008; Ferguson, Scheepers, & Sanford, 2010; Ferguson et al., 2008; Nieuwland & Martin, 2012; Stewart, Haigh, & Kidd, 2009; Urrutia, de Vega, & Bastiaansen, 2012). Ferguson and Sanford (2008) showed that despite a counterfactual context (e.g., “If cats were vegetarians”) sentences describing implausible real-world events (e.g., “Families would feed their cat a bowl of carrots”) incurred brief disruptions during reading (i.e., longer early fixations in eye-tracking; see also Ferguson et al., 2008, for N400 evidence), suggesting that counterfactual context came into play after real-world constraints had their effect (see Ferguson, Sanford, & Leuthold, 2007, on the role of negation). However, counterfactually consistent continuations had been rated as less plausible than real-world continuations (3.6 and 4.6, respectively, out of 5-‘highly plausible’). Recent eye-tracking results on comprehension of realistic counterfactuals (e.g., “If it had rained this morning Susan would have rushed to get to work. In the end, Susan arrived at work early (inconsistent)/late (consistent). . .”; Ferguson, 2012) suggest that readers do make factual inferences from counterfactuals (i.e., that it had not rained that morning and Susan had not rushed to work) yet maintain access to both counterfactual and factual interpretations (as evidenced by anomaly detection responses for counterfactual conditions compared to a factually consistent condition; see also de Vega et al., 2007; Santamaria, Espino, & Byrne, 2005).

Results from a recent ERP study by Nieuwland and Martin (2012) suggest that real-world knowledge does not modulate or delay the effect of propositional truth-value on processing counterfactuals about commonly-known historical events (e.g., “If NASA had not developed its Apollo Project, the first country to land on the moon would have been Russia/America surely”), as evidenced by reversed ERP patterns to the same lexical items as a function of the context. Perhaps that these alternative endings are easily computed because relevant information is part of our existing real-world knowledge (e.g., of the ‘Space Race’ between the USA and the USSR). Theories of counterfactual comprehension indeed assume that similarity between counterfactual worlds and the real world facilitates counterfactual reasoning (e.g., Byrne, 2007; Lewis, 1973; McCall, 1984). Thus, one unresolved issue is whether similar effects of propositional truth-value are observed when unrealistic consequences need to be computed on-the-fly. For the example sentence “If dogs had gills, Dobermans would breathe under water” the incremental interpretation involves the deduction that what holds for dogs holds for Dobermans (see Johnson-Laird, 1999), and an inductive inference regarding the potential consequences of a conceptual combination of or blend between ‘dog’ and ‘gills’ (e.g., Coulson & Fauconnier, 1999; Lakoff, 1987; Springer & Murphy, 1992; Turner & Fauconnier, 1998; see also Rips, 1975). Reasoning from this unrealistic conceptual combi-

nation within one sentence provides a strong test of the incremental contribution of counterfactual context.

The current hypotheses focused on the N400 (Kutas & Hillyard, 1980, 1984), an ERP waveform whose amplitude peaks at about 400 ms post-stimulus, with smaller amplitudes indexing facilitated retrieval from semantic memory as elicited by content words or other meaningful stimuli (for reviews see Kutas & Federmeier, 2011; Kutas, Van Petten, & Kluender, 2006). N400 effect onset, when ERP waveforms corresponding to different conditions start to diverge, is about 200–300 ms after visual word onset (e.g., Kutas & Hillyard, 1980, 1984; Van Berkum et al., 1999). With spoken words, N400 effects can start as early as 100–200 ms after word onset (e.g., Van Berkum, Zwitserlood, Hagoort, & Brown, 2003), after having heard only two or three phonemes and well before a word’s uniqueness point (e.g., Van Petten, Coulson, Rubin, Plante, & Parks, 1999). These well-established observations suggest that N400 effects that are elicited by words that are relatively unexpected (e.g., Kutas & Hillyard, 1984; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005) or that render sentences anomalous or false (e.g., Hagoort, Hald, Bastiaansen, & Petersson, 2004; Nieuwland & Kuperberg, 2008), reflect routine sense-making processes by which incoming words are related to the preceding context (e.g., Hagoort and Van Berkum (2007) and Kutas et al. (2006), for review). This involves retrieval from semantic memory as facilitated by linguistic and non-linguistic context and potentially intensified by attentional factors (see Federmeier & Laszlo, 2009; Kutas & Federmeier, 2000, 2011; Van Berkum, 2009; see also Baggio & Hagoort, 2011; Lau, Phillips, & Poeppel, 2008; Van Petten & Luka, 2006, for accounts of the neurobiology underlying N400 phenomena). In two separate experiments, the current study examined whether these processes are also sensitive to counterfactual context and propositional truth-value. In Experiment 1, participants read sentences that were anomalous with respect to real-world knowledge (e.g., “Dobermans would breathe under water”). In Experiment 2, new participants read these same sentences preceded by a counterfactual conditional premise that rendered the ‘locally anomalous’ sentence true (e.g., “If dogs had gills, Dobermans would breathe under water”).

The aim of Experiment 1 was to establish that without a counterfactual context, unrealistic consequences incurred semantic processing costs as indexed by the N400, similar to a real-world-false sentence compared to a true sentence (e.g., “Tuna breathe under poison/water”). This is not obvious due to potential processing differences related to verb tense between counterfactual real-world control sentences (conditional sentences had conditional verb tense, whereas real-world sentences did not). Conditional verb tense might mitigate the impact of propositional truth-value during processing, although earlier results suggest that this need not be the case (see Nieuwland & Martin, 2012).

Participants in Experiment 1 were native speakers of Spanish who read counterfactual-true/false control sentences and real-world-true/false control sentences (see Table 1), which were derived from the sentences from Experiment 2 (see Table 2). Critical words belonged to word pairs (e.g., ‘water’ and ‘poison’) that were matched

**Table 1**

Example sentences and translations with average truth-value rating per condition in Experiment 1.

Condition	Example sentences	Truth-value
Counterfactual-true control	<i>Los Dóberman respirarían bajo el <u>agua</u> sin problemas</i> <i>Una gota de sangre sería de color <u>verde</u> mayormente</i> <i>Los tomates podrían crear <u>adicción</u> fácilmente</i> Dobermans would breathe under <u>water</u> without problems A drop of blood would be the color <u>green</u> mainly Tomatoes could cause <u>addiction</u> easily	1.91 (0.80)
Counterfactual-false control	<i>Los Dóberman respirarían bajo el <u>veneno</u> sin problemas</i> <i>Una gota de sangre sería de color <u>blanco</u> mayormente</i> <i>Los tomates podrían crear <u>ilusión</u> fácilmente</i> Dobermans would breathe under <u>poison</u> without problems A drop of blood would be the color <u>white</u> mainly Tomatoes could cause <u>hope</u> easily	1.62 (0.70)
Real-world-true control	<i>Los atunes respiran bajo el <u>agua</u> sin problemas</i> <i>Una hoja de abeto es de color <u>verde</u> mayormente</i> <i>Los cigarros pueden crear <u>adicción</u> fácilmente</i> Tuna breathe under <u>water</u> without problems A fir tree's leaf is the color <u>green</u> mainly Cigarettes can cause <u>addiction</u> easily	5.80 (1.02)
Real-world-false control	<i>Los atunes respiran bajo el <u>veneno</u> sin problemas</i> <i>Una hoja de abeto es de color <u>blanco</u> mayormente</i> <i>Los cigarros pueden crear <u>ilusión</u> fácilmente</i> Tuna breathe under <u>poison</u> without problems A fir tree's leaf would be <u>white</u> mostly Cigarettes can cause <u>hope</u> easily	1.88 (0.85)
Filler control	<i>En Europa la mayoría de países usa el euro</i> <i>En España sólo el tabaco y el alcohol son legales</i> <i>La bicicleta no contamina en absoluto</i> En Europe the majority of countries uses the euro In Spain only tobacco and alcohol are legal Bicycles do not pollute at all	

Note: Standard deviations in parentheses. Critical words are underlined for expository purposes. For truth-value rating, 1 = False, 7 = True.

on relevant lexical variables (see Methods section). For the full counterfactual conditional sentences (Experiment 2), one word rendered both a counterfactual and a real-world sentence true and that the other word rendered them both false (e.g., Spanish equivalents of “If dogs had gills, Dobermans would breathe under water/poison without problems” and “Because fish have gills, tuna breathe under water/poison without problems”), as established in an independent truth-value rating test. Moreover, an independent cloze completion test was conducted to ensure that the ‘true’ words were roughly equally predictable from counterfactual contexts and real-world contexts, whereas ‘false’ words received zero cloze values.

The control sentences in Experiment 1 were constructed by omitting the first clause of the sentence, and another truth-value rating test was performed to ensure that participants would consider counterfactual control sentences and real-world-false control sentences as false, but real-world-true control sentences as true. The corresponding prediction was that compared to real-world-true control sentences, larger N400 responses would be observed for counterfactual-true/false control sentences and real-world-false control sentences.

A different pattern of N400 responses was predicted for corresponding counterfactual sentences in Experiment 2

but not for real-world sentences. If counterfactual consequences are computed incrementally such that propositional truth-value impacts semantic processing without delay, a comparable N400 effect of truth-value should be observed in real-world sentences and counterfactual sentences. In the abovementioned example, the incremental interpretation involves the deduction that what holds for dogs holds for Dobermans (for review, see Johnson-Laird, 1999), and an inductive inference regarding the type of niche that could be suitable for a dog with gills (e.g., Rips, 1975; for accounts of this process in terms of conceptual blending see Coulson & Fauconnier, 1999; Lakoff, 1987; Turner & Fauconnier, 1998). Alternatively, the gradual build-up of the counterfactual context may not eliminate interpretive problems posed by semantically anomalous consequences if build-up of the counterfactual context is somehow slower and impact of truth-value therefore delayed (e.g., Fischler, Bloom, Childers, Roucos, & Perry, 1983; Kounios & Holcomb, 1992; Urbach & Kutas, 2010), or because incoming statements are automatically mapped onto real-world knowledge despite the counterfactual context, in which case a smaller N400 effect of counterfactual truth-value is expected than of real-world truth-value. This would result from larger N400s to counterfactual-true sentences than real-world-true sentences, but a delayed

**Table 2**

Example sentences and translations with average truth-value rating and CW cloze value (%) in Experiment 2.

Condition	Example sentences	Truth-value	Cloze value
Counterfactual-true	<p><i>Si los perros tuviesen branquias, los Dóberman respirarían bajo el <u>agua</u> sin problemas</i>  <i>Si la sangre tuviese sobre todo clorofila, una gota de sangre sería de color <u>verde</u> mayormente</i>  <i>Si los vegetales tuvieran mucha nicotina, los tomates podrían crear <u>adicción</u> fácilmente</i>            If dogs had gills, Dobermans would breathe under <u>water</u> without problems            If blood had mostly chlorophyll, a drop of blood would be the color <u>green</u> mainly            If vegetables had a lot of nicotine, tomatoes could cause <u>addiction</u> easily</p>	5.86 (0.89)	0.65 (0.20)
Counterfactual-false	<p><i>Si los perros tuviesen branquias, los Dóberman respirarían bajo el <u>veneno</u> sin problemas.</i>  <i>Si la sangre tuviese sobre todo clorofila, una gota de sangre sería de color <u>blanco</u> mayormente.</i>  <i>Si los vegetales tuvieran mucha nicotina, los tomates podrían crear <u>ilusión</u> fácilmente</i>            If dogs had gills, Dobermans would breathe under <u>poison</u> without problems            If blood had mostly chlorophyll, a drop of blood would be the color <u>white</u> mainly            If vegetables had a lot of nicotine, tomatoes could cause <u>hope</u> easily.</p>	1.96 (0.83)	0
Real-world-true	<p><i>Como los peces tienen branquias, los atunes respiran bajo el <u>agua</u> sin problemas</i>  <i>Como la flora tiene sobre todo clorofila, una hoja de abeto es de color <u>verde</u> mayormente</i>  <i>Como el tabaco tiene mucha nicotina, los cigarros pueden crear <u>adicción</u> fácilmente</i>            Because fishes have gills, tuna breathe under <u>water</u> without problems.            Because flora has mostly chlorophyll, a fir tree's leaf is the color <u>green</u> mainly            Because tobacco has a lot of nicotine, cigarettes can cause <u>addiction</u> easily</p>	6.00 (0.94)	0.69 (0.22)
Real-world-false	<p><i>Como los peces tienen branquias, los atunes respiran bajo el <u>veneno</u> sin problemas</i>  <i>Como la flora tiene sobre todo clorofila, una hoja de abeto es de color <u>blanco</u> mayormente</i>  <i>Como el tabaco tiene mucha nicotina, los cigarros pueden crear <u>ilusión</u> fácilmente</i>            Because fishes have gills, tuna breathe under <u>poison</u> without problems            Because flora has mostly chlorophyll, a fir tree's leaf is the color <u>white</u> mainly            Because tobacco has a lot of nicotine, cigarettes can cause <u>hope</u> easily</p>	1.66 (0.71)	0
Filler	<p><i>Cada país tiene su moneda oficial, en Europa la mayoría de países usa el euro</i>  <i>La mayoría de las drogas son ilegales, en España sólo el tabaco y el alcohol son legales.</i>  <i>Hay transportes contaminantes y ecológicos, la bicicleta no contamina en absoluto.</i>            Every country has its official currency, in Europe the majority of countries uses the euro.            The majority of drugs are illegal, in Spain only tobacco and alcohol are legal.            There is polluting and ecological transportation, bicycles do not pollute at all.</p>		

Note: Standard deviations in parentheses. Critical words are underlined for expository purposes. For truth-value rating, 1 = False, 7 = True.

impact of counterfactual context might also entail smaller N400s to counterfactual-false sentences compared to real-world-false sentences.

## Experiment 1

### Methods

#### Development and pretest of materials for Experiment 1 and 2

150 Spanish sentence quadruplets were constructed that each consisted of two counterfactual sentences and two real-world sentences. Critical words were never sentence-final and could be nouns, verbs or predicates. Counterfactual-true sentences described the hypothetical consequences of a premise that was inconsistent with physical or biological facts of the world, involving characteristics of animate beings, inanimate objects, substances or events. In real-world-true sentences these animate beings or inanimate objects were replaced with alternative lexical items to make them consistent with real-world knowledge (e.g., “If dogs had gills, Dobermans would breathe under water” > “Because fish have gills, tuna breathe under water”). Counterfactuals started with the

conditional ‘Si’ and involved conditional verb tense, whereas real-world sentences started with ‘Como’ (‘since/because’) and did not contain conditional verb tense. False sentences were created by replacing critical words in true sentences with words that rendered them both false.

A sentence completion test was performed to establish the expectedness of true critical words. Twenty students of the University of the Basque Country completed one of two lists with one version of each sentence truncated before the critical word. They were instructed to complete sentences with the first sensible word coming to mind. Cloze value was computed as the percentage of participants who used the true critical word.

A subsequent rating pretest examined whether sentences (truncated after the critical word) were regarded as true or false. Twenty different students evaluated one of four lists of sentences so they saw only one condition per quadruplet, and decided for each sentence whether it was true (1 = False, 7 = True), skipping any they could not evaluate.

Based on these results, quadruplets were excluded when they had low cloze value (<.33), contained true/false sentences rated below/over 3.5, or containing sentences skipped by more than two participants. In the ultimate

set of 96 quadruplets (these can be found in the Supplementary Materials, along with the filler sentences), true and false sentences had relatively similar cloze values and truth-value ratings across conditions (see Table 2; truth-value ratings for the counterfactual-true and real-world-true sentences did not differ reliably ( $t(1,95) = 1.3$ , n.s.), but real-world-true cloze value was slightly higher than the counterfactual-true cloze value ( $t(1,95) = 2.08$ ,  $p < .05$ ), and counterfactual-false ratings were slightly higher than real-world-false ratings ( $t(1,95) = 3.52$ ,  $p < .05$ ). Critical words were matched for mean log frequency (True/False = 1.31/1.26; Davis & Perea, 2005) and word length (True/False = 6.7/6.4 letters, range 3–11 letters).

LSA semantic similarity values (SSV; see Landauer & Dumais, 1997) for the critical words were obtained using version 2 of the Gallito© software (<http://www.elsemanti-co.com>; Jorge-Botana, León, Olmos, & Hassan-Montero, 2010; Jorge-Botana, Olmos, & Barroso, 2012). This software uses a large training corpus to create representations of words and relationships between them within a multi-dimensional semantic space. Those representations are, importantly, insensitive to word order, syntax and overall propositional meaning. SSVs were obtained by cosine comparison of the critical word vector with the sentence vector (the sum of all word-vectors in the sentence) (counterfactual-true:  $M = .15$ ,  $SD = .12$ ; real-world-true:  $M = .18$ ,  $SD = .21$ ; counterfactual-false:  $M = .11$ ,  $SD = .10$ ; real-world-false:  $M = .10$ ,  $SD = .10$ ). Critical words in true sentences elicited higher similarity values than in false sentences ( $F_{1,84} = 17.74$ ,  $p = .001$ ,  $CI .058 \pm .016$ ), and there was a reliable truth-value by factuality interaction effect ( $F_{1,84} = 5.61$ ,  $p = .02$ ). Follow-up comparisons showed that the effect of truth-value was stronger in real-world sentences ( $F_{1,84} = 13.20$ ,  $p < .001$ ,  $CI .08 \pm .02$ ) than it was in counterfactual sentences ( $F_{1,84} = 5.94$ ,  $p < .05$ ,  $CI .036 \pm .015$ ). Moreover, real-world-true sentences elicited marginally higher values than counterfactual-true sentences ( $F_{1,84} = 3.7$ ,  $p = .058$ ,  $CI .037 \pm .019$ ).

Materials from Experiment 1 were based on those from Experiment 2 by omitting the first clause of the sentences and capitalizing the first letter of each second clause. A new pretest was performed to determine whether sentences (truncated after critical words) were regarded as true or false (see Table 1). Sixteen students who had not participated in other pretests evaluated one of four counterbalanced lists of sentences so they saw one condition per quadruplet, and decided for each sentence whether it was true (1 = False, 7 = True), skipping sentences they could not evaluate. No new cloze test was administered because 3 out of 4 conditions were on average rated as false, whereas the between-experiment comparison of real-world-true sentences (which may have differed in cloze value) per se was not relevant to the current hypothesis.

For the ERP experiment, four lists were created so that each of the 96 sentences appeared in only one condition per list, but in all conditions equally often across lists. Within each list, items were mixed with 120 filler sentences. The filler sentences were always consistent with real-world knowledge, and consisted of one clause (see Table 1).

### Participants

Twenty right-handed students (10 males; mean age = 21.2 years) gave written informed consent. All were native Spanish speakers, had no neurological or psychiatric disorders, nor had they participated in any of the pretests.

### Procedure

Participants read sentences from a monitor (black letters, light gray background), presented word-by-word (400 ms word duration, 200 ms inter-word-interval). Sentence-final words were followed by a fixation cross upon which participants could start the next sentence, or by a yes/no comprehension question that probed related knowledge. These 60 questions (30 requiring a 'yes' button-press response, average accuracy  $M = 81\%$  accuracy,  $SD = 9.1$ ) were distributed across sentence types (e.g., filler sentence: "The most famous comic books are the ones about superheroes", question: "Is Superman a superhero?"). Participants completed a practice-session and six break-separated experimental sessions. Total time-on-task was 40 min.

### Electroencephalogram recording, data processing and statistical analysis

The electroencephalogram (EEG) was recorded from 27 scalp electrodes (Fz, Cz, Pz, Fp1/2, F3/4, F7/8, FC1/2, FC5/6, C3/4, T7/8, CP1/2, CP5/6, P3/4, P7/8, O1/2, each referenced to the left mastoid; one additional right mastoid electrode and four additional electro-oculogram electrodes), amplified (band-pass filtered at 0.01–30 Hz), and digitized at 250 Hz. Impedance was kept below 5 k $\Omega$ . After re-referencing to the average of the left and right mastoid electrodes, ocular artifacts were corrected using a procedure based on independent component analysis, and single-trial waveforms were screened for artifacts during 1000 ms epochs (starting 150 ms before critical word onset). Four participants were excluded due to excessive artifacts (trial loss > 40%). For the remaining 16 participants, average ERPs (normalized by subtraction to a 150-ms pre-stimulus baseline) were computed over artifact-free trials per condition (average percentage of included trials = 93%, range = 88–97% across the four conditions).

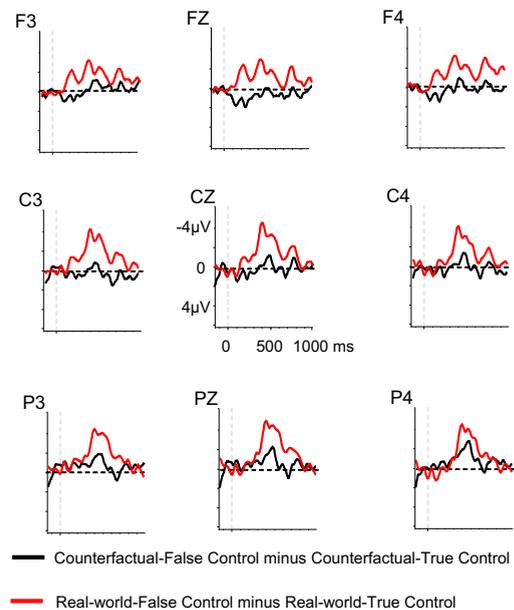
First, using average amplitude per condition across all EEG electrodes a 2(factuality: counterfactual, real-world)  $\times$  2(truth-value: true, false) repeated measures analysis of variance (ANOVA) was performed in consecutive 100 ms time windows between 250 and 650 ms after critical word onset. All confidence intervals are at an alpha level of 0.05. Then, additional analyses were performed to explore the scalp distribution of the observed effects and to reveal potential scalp distribution differences between observed effects in counterfactual sentences and real-world sentences. Electrodes were grouped into quadrants according to hemisphere (left or right of the midline) and according to whether they were anterior or posterior to the crossline, thus excluding midline and crossline electrodes (left-anterior: FP1, F3, F7, FC1, FC5; right-anterior: FP2, F4, F8, FC2, FC6; left-posterior: CP1, CP5, P3, P7, O1; right-posterior: CP2, CP6, P4, P8, O2). For each of the time windows, a 2(factuality: counterfactual control, real-world control)  $\times$  2(truth-value: true control, false

control)  $\times$  2(anteriority: anterior, posterior)  $\times$  2(hemisphere: left, right) ANOVA was performed.

## Results

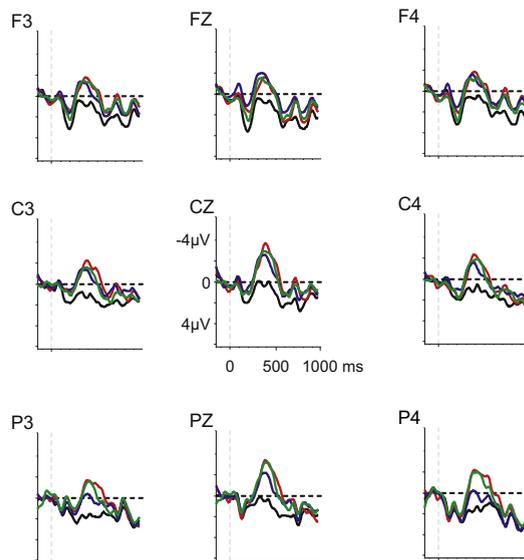
Critical words elicited more negative N400s in counterfactual-false control sentences and real-world-false control sentences compared to real-world-true control sentences (see Fig. 1), visible at most electrodes. Importantly, counterfactual-true control sentences elicited N400s that followed those elicited by false control sentences at most electrodes, but elicited only slightly increased N400s compared to real-world-true control sentences at posterior electrodes (see Fig. 2).

The repeated measures ANOVAs confirmed that counterfactual control sentences elicited more negative ERPs than real-world control sentences between 250 and 350 ms ( $F_{1,15} = 6.21, p < .05, CI -0.85 \pm .34$ ), between 350 and 450 ms ( $F_{1,15} = 4.97, p < .05, CI -0.94 \pm .42$ ). In addition, false control sentences overall elicited larger N400s than true control sentences between 350 and 450 ms ( $F_{1,15} = 19.81, p < .001, CI -1.63 \pm .37$ ), between 450 and 550 ms ( $F_{1,15} = 25.76, p < .001, CI -1.57 \pm .31$ ), and between 550 and 650 ms ( $F_{1,15} = 4.78, p < .05, CI -0.71 \pm .33$ ). However, this true-false effect differed reliably between counterfactual and real-world sentences in the 350 and 450 ms time window ( $F_{1,15} = 8.12, p < .012$ ) and marginally between 450 and 550 ms ( $F_{1,15} = 3.68, p < .01$ ). Follow-up pairwise comparisons (LSD) revealed that between 350



**Fig. 2.** Difference waves for Experiment 1 (counterfactual-false control minus counterfactual-true control sentences, real-world-false control minus real-world-true control sentences).

and 450 ms and between 450 and 550 ms, counterfactual-true control sentences elicited more negative ERPs than real-world-true control sentences (350–450:



Counterfactual-True Control: Dobermans would breathe under water without problems.

Counterfactual-False Control: Dobermans would breathe under poison without problems.

Real-world-True Control: Tuna breathe under water without problems.

Real-world-False Control: Tuna breathe under poison without problems.

**Fig. 1.** Electrophysiological effects of truth-value in counterfactual control and real-world control sentences in Experiment 1. The waveforms show the grand average event-related potentials elicited by critical words per condition at nine selected electrodes. In this and all following graphs, negative voltage is plotted upwards and waveforms are filtered (5 Hz high cut-off, 12 dB/oct) for presentation purpose only. Stimuli consisted of counterfactual and real-world Spanish sentences that were either true or false (translated examples are provided below the graphs with corresponding condition labels). Note that counterfactual-true-control sentences are not considered true but this condition label is used to facilitate the between-experiment comparison.

$F_{1,15} = 19.32$ ,  $p = .001$ , CI  $-1.89 \pm .43$ ; 450–550:  $F_{1,15} = 14.43$ ,  $p < .01$ , CI  $-1.25 \pm .33$ ), and real-world-false control sentences elicited more negative ERPs than real-world-true control sentences (350–450:  $F_{1,15} = 23.38$ ,  $p < .001$ , CI  $-2.58 \pm .53$ ; 450–550:  $F_{1,15} = 21.75$ ,  $p < .001$ , CI  $-2.24 \pm .48$ ).

#### Scalp distribution analyses

Between 350 and 450 ms, a significant truth-value control by anteriority interaction effect was observed ( $F_{1,15} = 7.55$ ,  $p < .05$ ): false control sentences elicited more negative ERPs compared to true control sentences at posterior electrodes ( $F_{1,15} = 27.65$ ,  $p < .001$ , CI  $-2.06 \pm .39$ ) than they did at anterior electrodes ( $F_{1,15} = 9.10$ ,  $p < .01$ , CI  $-1.2 \pm .39$ ).

Between 450 and 550 ms, a significant factuality by truth-value control by anteriority by hemisphere 4-way interaction was observed ( $F_{1,15} = 4.82$ ,  $p < .05$ ): a statistically reliable truth-value control by factuality interaction effect was only observed at left-posterior channels ( $F_{1,15} = 6.46$ ,  $p < .05$ ) and right-posterior channels ( $F_{1,15} = 4.78$ ,  $p < .05$ ). At left-posterior and right-posterior channels, counterfactual-true control sentences elicited more negative ERPs than real-world-true control sentences (left:  $F_{1,15} = 17.27$ ,  $p = .001$ , CI  $-1.22 \pm .29$ ; right:  $F_{1,15} = 12.72$ ,  $p < .01$ , CI  $-1.20 \pm .34$ ) and real-world-false control sentences elicited more negative ERPs than real-world-true control sentences (left  $F_{1,15} = 35.55$ ,  $p < .001$ , CI  $-2.81 \pm .47$ ; right  $F_{1,15} = 26.46$ ,  $p < .001$ , CI  $-2.93 \pm .57$ ). Only at right-posterior channels, however, did counterfactual-false control sentences elicit more negative ERPs than counterfactual-true control sentences ( $F_{1,15} = 14.94$ ,  $p < .01$ , CI  $-1.49 \pm .39$ ).

#### Discussion

As predicted, counterfactual-true control sentences elicited N400s that were larger than those in real-world-true control sentences, and that were comparable to N400s to the false control sentences. These results therefore established that an N400 effect of truth-value could be obtained despite the conditional verb tense. It must be noted that counterfactual-true control sentences elicited N400 responses that were less pronounced at parietal-occipital electrodes than N400 responses to the false control sentences, and the difference between counterfactual-true control and real-world-true control sentences was therefore more centrally distributed than is common for N400 modulations (e.g., Kutas et al., 2006). It is unclear why this difference occurred, because frontally distributed N400 effects are not generally reported for sentence reading paradigms, although sometimes for spoken sentence or picture comprehension (e.g., Kutas & Federmeier, 2011; Kutas et al., 2006).

#### Experiment 2

As outlined in the Introduction section in more detail, the aim of Experiment 2 was to show an effect of propositional truth-value in unrealistic counterfactual conditional

sentences. Effectively, this means that the supportive context offered by the counterfactual premise should preclude the larger N400 responses that were seen to counterfactual-true control sentences compared to real-world-true sentences in Experiment 1. Moreover, no ERP differences were predicted for counterfactual-false and real-world-false sentences, replicating the observed pattern to the corresponding control sentences in Experiment 1.

#### Methods

##### Stimulus materials

Four lists were created so that each of the 96 full sentences appeared in only one condition per list, but in all conditions equally often across lists. Within each list, items were mixed with 120 filler sentences. The filler sentences were always consistent with real-world knowledge, did not start with 'Si' or 'Como', and consisted of two clauses separated by a comma (see examples in Table 2).

##### Participants

20 Right-handed students (11 males; mean age = 20.9 years) gave written informed consent. All were native Spanish speakers, had no neurological or psychiatric disorders, nor had they participated in any of the pretests or in Experiment 1.

##### Procedure

The first clause was presented for 4000 ms, during which participants were allowed to blink or move, followed by a blank screen for 500 ms. The duration of the first clause was based on the average number of words in that clause (6.9 words), such that whole-clause presentation would – on average – be faster than word by word presentation while taking into account that participants often needed to read the clause across two lines. The second clause was presented word-by-word (400 ms word duration, 200 ms inter-word-interval). Sentence-final words were followed by a fixation cross upon which participants could start the next sentence or by a yes/no simple comprehension question that probed world knowledge related to the preceding sentence. These 60 questions were included to keep participants more alert during the experiment (30 requiring a 'yes' button-press response, average accuracy  $M = 82\%$  accuracy,  $SD = 9.7$ ), and were distributed across sentence types (e.g., filler sentence: "Many people like to read comic books, the most famous comic books are the ones about superheroes", question: "Is Superman a superhero?"). Participants completed a practice-session and six break-separated experimental sessions. Total time-on-task was approximately 50 min.

##### Electroencephalogram recording, data processing and statistical analysis

Recording, processing and analysis of the EEG data occurred identically to in Experiment 1. 2 participants were excluded due to excessive artifacts (trial loss > 40%). For the remaining 18 participants, average ERPs (normalized by subtraction to a 150-ms pre-stimulus baseline) were computed over artifact-free trials per condition

(average percentage of trials = 96%, range = 95–97% across conditions).

The statistical analyses that were performed were similar to those reported for Experiment 1, but the variable label 'truth-value control' was changed into 'truth-value'. In addition,  $2(\text{truth-value: true, false}) \times 2(\text{factuality: counterfactual, real-world}) \times 2(\text{experiment: experiment 1, experiment 2}) \times 2(\text{anteriority: anterior, posterior}) \times 2(\text{hemisphere: left, right})$  repeated measures ANOVAs were performed to test for statistical differences between Experiment 1 and Experiment 2 for each time window. Greenhouse–Geisser correction was used for  $F$  tests with 2 or more degrees of freedom. Additional  $2(\text{truth-value: true, false}) \times 2(\text{factuality: counterfactual, real-world}) \times 2(\text{experiment: experiment 1, experiment 2})$  ANOVAs were performed using the average voltage per condition at midline electrodes (FZ, CZ, Pz) and at crossline electrodes (T7/8, C3/4).

## Results

Critical words elicited larger (more negative) N400s in the counterfactual-false and real-world-false sentences compared to counterfactual-true and real-world-true sentences (see Fig. 3), this pattern is visible at all electrodes. These N400 effects of truth-value in counterfactual and real-world sentences started at about 200 ms after critical word onset, dissipated before 800 ms, and had a broad central distribution (see Fig. 4). Statistical analyses using all electrodes confirmed that critical words in false sentences elicited more negative ERPs than true sentences between 250 and 350 ms ( $F_{1,17} = 34.97, p < .001, CI -1.5 \pm .27$ ), between 350 and 450 ms ( $F_{1,17} = 82.24, p < .001, CI -3.4 \pm .38$ ), between 450 and 550 ms ( $F_{1,17} = 58.37, p < .001, CI -3.34 \pm .44$ ), and between 550 and 650 ms

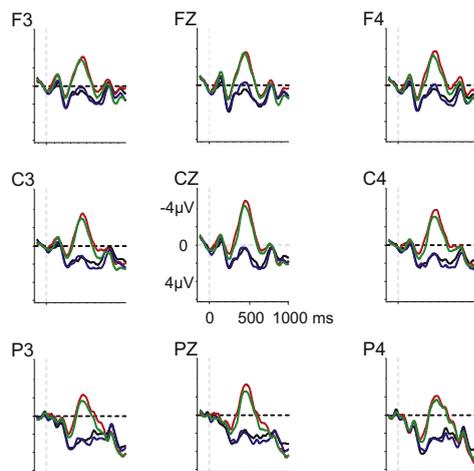
( $F_{1,17} = 20.81, p < .001, CI -1.50 \pm .33$ ). Crucially, there was no significant main effect of factuality ( $F < 1$  for factuality in all time windows), and the effect of truth-value did not differ for counterfactual and real-world sentences in any time window ( $F < 1$  for all truth-value by factuality interactions).

### Scalp distribution analyses

Between 350 and 550 ms, false sentences elicited more negative ERPs compared to true sentences at posterior electrodes (350–450:  $F_{1,17} = 93.31, p < .001, CI -3.63 \pm .38$ ; 450–550:  $F_{1,17} = 91.69, p < .001, CI -3.80 \pm .40$ ) than they did at anterior electrodes (350–450:  $F_{1,17} = 53.30, p < .001, CI -2.97 \pm .41$ ; 450–550:  $F_{1,17} = 27.17, p < .001, CI -2.70 \pm .52$ ). Between 550 and 650 ms, false sentences elicited more negative ERPs compared to true sentences at right-hemisphere electrodes ( $F_{1,17} = 28.71, p < .001, CI -1.72 \pm .32$ ) than they did at left-hemisphere electrodes ( $F_{1,17} = 11.24, p < .01, CI -1.20 \pm .36$ ).

### Between-experiment analyses

The main purpose of the between-experiment analyses was to establish that the experiments yielded a reliably different pattern of results for the interaction between truth-value and factuality. Using all electrodes, the truth-value by factuality by experiment 3-way interaction was marginally significant between 350 and 450 ms ( $F_{1,32} = 3.23, p < .1$ ). In this time window, the truth-value by factuality by anteriority by experiment 4-way interaction was fully significant ( $F(1,32) = 7.94, p < .01$ ). Follow-up revealed that the truth-value by factuality by experiment 3-way interaction was only statistically significant at anterior electrodes ( $F(1,32) = 4.56, p < .05$ ), consistent with the observation that the between-experiment



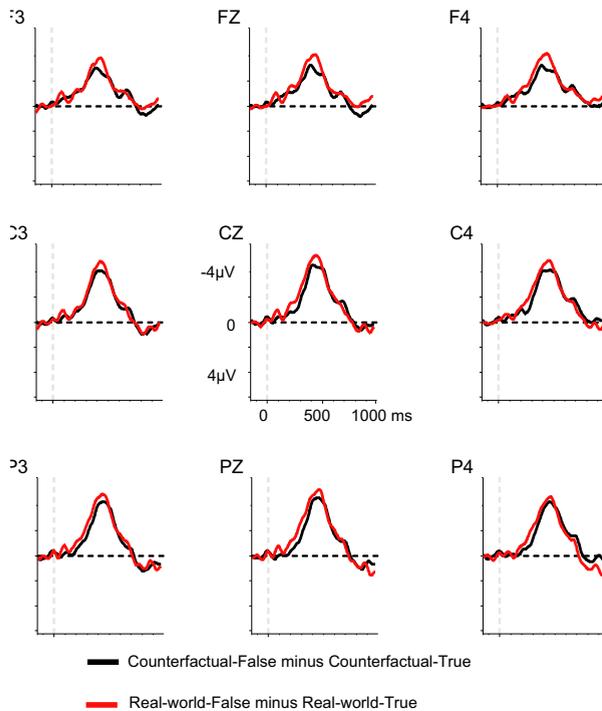
Counterfactual-True: If dogs had gills, Dobermans would breathe under water without problems.

Counterfactual-False: If dogs had gills, Dobermans would breathe under poison without problems.

Real-world-True: Because fish have gills, tuna breathe under water without problems.

Real-world-False: Because fish have gills, tuna breathe under poison without problems.

Fig. 3. Grand average event-related potentials elicited by critical words per condition in Experiment 2.



**Fig. 4.** Difference waves for Experiment 2 (counterfactual-false minus counterfactual-true sentences, real-world-false minus real-world-true sentences).

differences in N400 modulations across conditions were most pronounced at anterior electrodes (reflecting the stronger interaction for true/false control and factuality at anterior channels in Experiment 1).

Between 450 and 550 ms, the truth-value by factuality by anteriority by hemisphere by experiment 5-way interaction was fully significant ( $F(1,32) = 4.34, p < .05$ ), consistent with the earlier observations of a significant factuality by truth-value control by anteriority by hemisphere 4-way interaction in Experiment 1 but not in Experiment 2.

#### Additional analyses of cloze probability

Further analyses focused on the influence of cloze probability on N400 activity (350–450 ms time window) elicited by the counterfactual-true and real-world-true sentences. These analyses aimed to show that a comparison of counterfactual-true and real-world-true sentences would (1) yield similar results of factuality in sentences with high cloze probability and in sentences with moderate cloze probability when cloze probability was matched between factual and counterfactual conditions, but (2) yield an overall larger N400 for moderate cloze sentences as compared to high cloze sentences. This was accomplished by dividing the item-set in two subsets based on an approximate median-split of the average cloze probability for counterfactual and real-world sentences: one set had a relatively high average cloze probability ( $M = .83, SD = .09$ ) and the other set had a relatively moderate average cloze probability ( $M = .49, SD = .10$ ). The two sets of items were matched in terms of critical word length

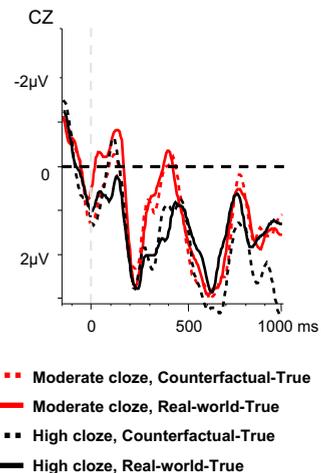
and log frequency, as well as on the truth-value ratings of the counterfactual-true and real-world-true sentences (all  $F_s < 2.2, n.s.$ ).

The results for a representative electrode Cz are shown in Fig. 5. For the statistical analysis, averages were computed over the more centrally located electrodes that are most sensitive to N400 activity (F3/F4, C3/C4, P3/P4, Fz, Cz, Pz, FC1/2, CP1/2, FC5/6, CP5/6). A 2(factuality: counterfactual, real-world)  $\times$  2(average cloze: high, moderate) ANOVA confirmed that critical words in low cloze sentences elicited more negative ERPs in the N400 time window than in high cloze sentences ( $F_{1,17} = 4.51, p < .05, CI -1.29 \pm .55$ ), and that this pattern was similar for counterfactual and real-world sentences.

#### Discussion

This ERP experiment examined comprehension of counterfactual and real-world sentences varying in truth-value. N400 effects of propositional truth-value were obtained in counterfactual and real-world sentences, and these effects had the typical central-posterior, slightly right-lateralized N400 effect distribution (e.g., Kutas et al., 2006). Importantly, ERPs for counterfactual-true and real-world-true sentences fully overlapped (as did the counterfactual-false and real-world-false sentences) and associated N400 effects were statistically indistinguishable.

Of note, the absence of an N400 difference between counterfactual-true and real-world-true sentences is unlikely to be the result of lexical-associative priming. Lexical co-occurrence analysis showed that the critical words had greater semantic similarity to the real-world context than to the counterfactual context, while critical words had greater similarity to true contexts than to false contexts. This reflects that, on average, words like ‘water’ are more similar to the combination of ‘fish’-‘tuna’-‘gills’-‘breathe’ than to the combination of ‘dogs’-‘gills’-‘Dobermanns’-‘breathe’, whereas ‘poison’ would be much less similar to these combinations. This makes sense in terms of the



**Fig. 5.** ERP waveforms at Cz for high-cloze and moderate-cloze counterfactual-true and real-world-true sentences in Experiment 2.

additivity assumption in spreading activation models of semantic memory (e.g., Collins & Loftus, 1975) and the literature on summation priming in word lists (e.g., Balota & Paul, 1996). The current ERP results thus suggest that the N400 is not solely sensitive to such lexical priming but can reflect interpretation at a message-level representation of the discourse (e.g., Camblin, Gordon, & Swaab, 2007; Camblin, Ledoux, Boudewyn, Gordon, & Swaab, 2007; Nieuwland & Kuperberg, 2008; Nieuwland & Martin, 2012; Otten & Van Berkum, 2007; see also Traxler, Foss, Seely, Kaup, & Morris, 2000).

These results, in particular the fact that counterfactual-true sentences did not incur any visible semantic processing costs upon encountering the critical words over real-world-true sentences, suggest that the counterfactual context completely eliminated the interpretive problems posed by the semantically anomalous consequence, and that incoming words can therefore be mapped without any delay onto the interpretive context.

### General discussion

Two ERP experiments examined electrophysiological responses to propositional truth-value of sentences about biologically or physically unrealistic counterfactual worlds (e.g., Spanish equivalents of “If dogs had gills, Dobermans would breathe under water/poison”, true/false) or of real-world sentences (“Because fish have gills, tuna breathe under water/poison”, true/false). Experiment 1 established that without the counterfactual premise (e.g., “Dobermans would breathe under water/poison”, false/false, versus “Tuna breathe under water/poison”, true/false), each type of false sentence elicited larger N400 responses than true sentences. In Experiment 2, participants read the counterfactual or real-world-true/false sentences with the supportive context. Counterfactual-false sentences and real-world-false sentences elicited similar N400 responses, which were reliably larger than their counterpart true sentences in both contexts, suggesting that incorporation of the counterfactual context eliminated the interpretive problems posed by locally anomalous sentences.

The results suggest that, even in sentences that describe impossible events, if counterfactual consequences are sufficiently plausible given the premise then ongoing semantic processing is impacted by propositional truth-value without an observable delay or disruption. This is inconsistent with accounts of discourse comprehension that assume a context-free level of semantic analysis based on local lexical-semantic associations that precedes (whether in a cascaded or in a pure serial manner) a phase in which implications for the wider context are computed (e.g. Kintsch, 1988; Myers & O'Brien, 1998). The results are also inconsistent with accounts of language comprehension that assume a temporal delay in contextual effects specifically in sentences with logical operators such as negation and scalar quantifiers (e.g., Fischler et al., 1983; Kounios & Holcomb, 1992) and counterfactuals (e.g., Carpenter, 1973; Carpenter & Just, 1975). Instead, the results are consistent with theories of language comprehension that do not assume a delayed contribution of discourse context (see

Cook & Myers, 2004). The results provide further evidence that people effortlessly map incoming utterances onto what they think is true and what they consider relevant (e.g., Hagoort et al., 2004; Nieuwland & Kuperberg, 2008), and may reflect how contextual constraints guide expectations about upcoming input (e.g., DeLong, Urbach, & Kutas, 2005; Federmeier, 2007; Kutas, DeLong, & Smith, 2011; Van Berkum, 2009).

The immediate impact of propositional truth-value in counterfactual sentences about unrealistic worlds, as reported here, contrasts with several reports in the literature that pitted contextual relevance with real-world truth-value (e.g., Hald et al., 2007) or real-world possibility (e.g., Ferguson & Sanford, 2008; Ferguson et al., 2008; Warren et al., 2008). As noted in the Introduction, these other reports may have involved a context that rendered an impossible consequence mildly congruent, yet still less predictable or plausible than a *possible* consequence. It has long been established that the N400 is strongly correlated to predictability, with its amplitude being inversely correlated to cloze probability (e.g., DeLong et al., 2005; Kutas & Hillyard, 1984; see also Federmeier & Kutas, 1999; Otten & Van Berkum, 2008; Van Berkum et al., 2005). A post hoc analysis was conducted to disentangle effects of factuality from those of predictability, by comparing N400 effects in counterfactual and real-world sentences with a relatively high cloze value to those in sentences with a relatively moderate cloze value: whereas the N400 was reliably modulated by cloze value (larger N400 for moderate cloze sentences), the effect of factuality was not affected by cloze value (reflecting the matching of cloze value for counterfactual and real-world sentences within each subset). Thus, the observed pattern seems not to be contingent on high predictability of the critical words, and future research could further elucidate this issue by examining the comprehension of counterfactual sentences with low-predictable critical words that are rated as equally plausible or true compared to a real-world-true control (although these may be difficult to construct: see Frisson, Rayner, & Pickering, 2005; Matsuki et al., 2011, for discussion). In contrast, earlier work has shown that N400 effects of truth-value are not observed in literally true but pragmatically infelicitous sentences that, effectively, are zero-cloze. At the very least, the current results are most probably contingent on having counterfactual and real-world sentences that are matched on contextual constraint leading up to critical words, and on the plausibility or truth-value as rendered by the critical words. In this study, truth-value ratings are effectively similar to plausibility<sup>2</sup> ratings (“If X, how plausible is Y?”).

The present studies addressed the impact of counterfactual truth-value at the critical words, but this only represents a snapshot of counterfactual processing, and could not tap directly into the construction of the counterfactual

<sup>2</sup> However, one can obtain plausibility ratings for sentences of which truth-value cannot be established, e.g., “He spread the warm bread with socks”), whereas it would be infelicitous to ask for the plausibility (rather than truth-value) of straightforwardly false sentences (e.g., “The capital of France is London”) unless participants are instructed to rate how plausible it is that somebody would utter this sentence.

world itself. Importantly, the claim of this paper is thus not that comprehension of counterfactual sentences is functionally identical to or “as easy” as with regular sentences, but, rather, that the implications of the counterfactual context and real-world context have both been computed such that propositional truth-value directly impacted online semantic processing at the critical word. It is thus very well possible, and perhaps plausible, that the setting-up itself of a counterfactual context is more costly than that of a real-world context. Additionally, the current study does not have straightforward implications for theories of counterfactual reasoning, which are concerned with the balance between (and activation patterns of) counterfactual and factual representations (e.g., Byrne, 2002; Evans, 2006). It remains an open and important question how contextual constraints build-up as participants process language, and how readers construe the counterfactual world to serve as a context for interpretation. In the following section I will briefly describe what this could entail.

#### *Constructing a counterfactual context*

An initial interpretation of a counterfactual premise might follow the particular constraints of the conceptual combination that is required (e.g., “If blood contained chlorophyll”; see Costello & Keane, 2000; Springer & Murphy, 1992; for accounts in terms of conceptual blending, see Coulson, 2001; Coulson & Fauconnier, 1999; Coulson & Oakley, 2005). Readers might not consider information that undermines coherency of the counterfactual world (e.g., that high levels of chlorophyll in blood are potentially deadly to animals and humans, and that blood would not be blood if it contained mostly chlorophyll, see Murphy & Medin, 1985), as part of the ‘suspension of disbelief’ required for such sentences (e.g., Searle, 1975). Readers might instead focus on the most diagnostic and plausible changes by the counterfactual scenario as compared to the real-world. For example, because chlorophyll is strongly associated with photosynthesis as well as with being that what makes plant green, it is possible that the counterfactual premise leads readers to think about one or both of the following things: animals can now do photosynthesis, and blood has the color green. Guided by the unfolding sentence and presentation of “the color”, readers can narrow down to the most relevant consequences. Perhaps that a dedicated follow-up study could reveal processing differences already before any influence of propositional truth-value would be expected (e.g., at the word ‘color’).

Counterfactual sentence comprehension thus seem to involve online reasoning and inferencing. In the current items, counterfactual consequences were about a noun phrase that was never identical to the noun phrase introduced in the premise but always constituted a subset thereof (e.g., ‘Dobermans’ following ‘dogs’), requiring a deductive inference (what is true for dogs is true for Dobermans) and inductive inference (likelihood that dogs with gills would indeed breathe under water). The latter inference might, similar to reasoning about the premise, require suspension of disbelief: people might consider that having gills is predictive of living under water (see Murphy

& Ross, 2010; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990), whereas because dogs are not great swimmers they might not survive under water and would nevertheless be better off living as land-animals with gills (having gills does not mean they cannot also have lungs). This does not necessarily mean that participants actively suppress these kinds of thoughts; they might simply not become aware of inconsistencies as they read, especially if they are not informative or relevant to the unfolding discourse (see also Evans, 2006). The current pattern of results might not hold true, however, for individuals who apply real-world knowledge more rigidly during language comprehension (e.g., Morsanyi & Handley, 2012; Scott, Baron-Cohen, & Leslie, 1999).

#### **Conclusion**

People are able to establish whether a sentence is hypothetically true despite knowledge that what it describes can never be true given the biological or physical laws of the world. The present study examined brain responses to propositional truth-value of counterfactual conditional Spanish sentences about such unrealistic worlds. In contrast to earlier findings regarding comprehension of sentences with logical operators (e.g., Fischler et al., 1983) and comprehension of counterfactual language (e.g., Ferguson et al., 2008; Warren et al., 2008), the present results show that when described consequences are sufficiently plausible given an unrealistic counterfactual premise, ongoing semantic processing is directly impacted by propositional truth-value.

#### **Acknowledgments**

I wish to thank Eneko Antón for help with stimulus construction and data collection, and Andrea Martin for comments on an earlier version of this manuscript. This work is supported by the Spanish Ministry of Science and Innovation through a Plan National research grant and a Ramón y Cajal fellowship.

#### **A. Supplementary material**

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jml.2012.08.003>.

#### **References**

- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.
- Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, *26*, 1338–1367.
- Balota, D. A., & Paul, S. (1996). Summation of activation: Evidence from multiple primes than converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 827–845.
- Byrne, R. M. J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences*, *6*, 426–431.
- Byrne, R. M. J. (2007). Précis of the rational imagination: How people create alternatives to reality. *Behavioral and Brain Sciences*, *30*, 439–480.

- Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, 56, 103–128.
- Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research*, 1146, 172–184.
- Carpenter, P. A. (1973). Extracting information from counterfactual clauses. *Journal of Verbal Learning and Verbal Behavior*, 12, 512–521.
- Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82, 45–73.
- Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82, 407–428.
- Cook, A., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: The influences of context and world knowledge. *Journal of Memory and Language*, 50, 268–288.
- Cosmides, L., & Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentations* (pp. 53–115). Oxford: Oxford University Press.
- Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, 24, 299–349.
- Coulson, S. (2001). *Semantic leaps: Frame-shifting and conceptual blending in meaning construction*. Cambridge: Cambridge University Press.
- Coulson, S., & Fauconnier, G. (1999). Fake guns and stone lions: Conceptual blending and privative adjectives. In B. Fox, D. Jurafsky, & L. Michaelis (Eds.), *Cognition and Function in Language*. Palo Alto, CA: CSLI.
- Coulson, S., & Oakley, T. (2005). Blending and coded meaning: Literal and figurative meanings in cognitive semantics. *Journal of Pragmatics*, 37, 1510–1536.
- Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research*, 1155, 147–162.
- Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37, 665–671.
- de Vega, M., Urrutia, M., & Rizzo, R. (2007). Cancelling updating in the comprehension of counterfactuals embedded in narratives. *Memory & Cognition*, 35, 1410–1421.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117–1121.
- Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Co-Registration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, 140, 552–572.
- Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, 32, 551–559.
- Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13, 378–395.
- Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. New York, NY, US: Basic Books Inc.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44, 491–505.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41, 469–495.
- Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 51, pp. 1–44). Burlington: Academic Press.
- Ferguson, H. J. (2012). Eye movements reveal rapid concurrent access to factual and counterfactual interpretations of the world. *Quarterly Journal of Experimental Psychology*, 65, 939–961.
- Ferguson, H. J., Sanford, A. J., & Leuthold, H. (2007). Real-world interference in detecting violations of counterfactual and negated worlds. In: *Poster presented at cognitive neuroscience society annual meeting*. New York, NY.
- Ferguson, H. J., & Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language*, 58, 609–626.
- Ferguson, H. J., Sanford, A. J., & Leuthold, H. (2008). Eye-movements and ERPs reveal the time-course of processing negation and remitting counterfactual worlds. *Brain Research*, 1236, 113–125.
- Ferguson, H. J., Scheepers, C., & Sanford, A. J. (2010). Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes*, 25, 297–346.
- Filcik, R. (2008). Contextual override of pragmatic anomalies: Evidence from eye movements. *Cognition*, 106, 1038–1046.
- Filcik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. *Psychophysiology*, 45, 554–558.
- Fischler, I., Bloom, P., Childers, D., Roucos, S., & Perry, N. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20, 400–409.
- Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 862–877.
- Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language*, 42, 526–544.
- Gerrig, R. J., & Bortfeld, H. (1999). Sense creation in and out of discourse contexts. *Journal of Memory and Language*, 41, 457–468.
- Gerrig, R. J., & Murphy, G. L. (1992). Contextual influences on the comprehension of complex concepts. *Language and Cognitive Processes*, 7, 205–230.
- Gerrig, R. J., & O'Brien, E. J. (2005). The scope of memory-based processing. *Discourse Processes*, 39, 225–242.
- Gibbs, R. (1979). Contextual effects in understanding indirect requests. *Discourse Processes*, 2, 1–10.
- Hagoort, P., Hald, L. A., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438–441.
- Hagoort, P., & van Berkum, J. J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society, Series B*, 362, 801–811.
- Hald, L. A., Steenbeek-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research*, 1146, 210–218.
- Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, 124, 62–82.
- Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology*, 50, 109–135.
- Jorge-Botana, G., Olmos, R., & Barroso, A. (2012). Gallito (Version 2.0.1) [NLP Software]. Retrieved from <http://www.elsemantic.com/descargas.html>.
- Jorge-Botana, G., León, J. A., Olmos, R., & Hassan-Montero, Y. (2010). Visualizing polysemy using LSA and the predication algorithm. *Journal of the American Society for Information Science and Technology*, 61, 1706–1724.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163–182.
- Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General*, 121, 459–479.
- Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, 4, 463–470.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.
- Kutas, M., Van Petten, C., & Klueder, R. (2006). Psycholinguistics electrified II: 1994–2005. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659–724). New York: Elsevier.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL, US: University of Chicago Press.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De) constructing the N400. *Nature Reviews Neuroscience*, 9, 920–933.

- Lewis, D. (1973). *Counterfactuals*. Cambridge, Mass.: Harvard University Press.
- Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science*, 189, 226–228.
- Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 37, 913–934.
- McCall, S. (1984). Counterfactuals based on real possible worlds. *Nous*, 18, 462–477.
- McMullen, M. N., & Markman, K. D. (2002). Affective impact of close counterfactuals: Implications of possible futures for possible pasts. *Journal of Experimental Social Psychology*, 38, 64–70.
- Morsanyi, K., & Handley, S. J. (2012). Reasoning on the basis of fantasy content: Two studies with high-functioning autistic adolescents. *Journal of Autism and Developmental Disorders*. <http://dx.doi.org/10.1007/s10803-012-1477-0>.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- Murphy, G. L., & Ross, B. H. (2010). Category vs. object knowledge in category-based induction. *Journal of Memory and Language*, 63, 1–17.
- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26, 131–157.
- Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, 19, 1213–1218.
- Nieuwland, M. S., & Martin, A. E. (2012). If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition*, 122, 102–109.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18, 1098–1111.
- Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, 97, 185–200.
- Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research*, 1153, 166–177.
- Otten, M., & Van Berkum, J. J. A. (2008). Discourse-based lexical anticipation: Prediction or priming? *Discourse Processes*, 45, 464–496.
- Rayner, K., Cook, A. E., Juhasz, B. J., & Frazier, L. (2006). Immediate disambiguation of lexically ambiguous words during reading: Evidence from eye movements. *British Journal of Psychology*, 97, 467–482.
- Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, 14, 665–681.
- Santamaria, C., Espino, O., & Byrne, R. M. J. (2005). Counterfactual and semifactual conditionals prime alternative possibilities. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31, 1149–1154.
- Scott, F. J., Baron-Cohen, S., & Leslie, A. (1999). 'If pigs could fly': A test of counterfactual reasoning and pretence in children with autism. *British Journal of Developmental Psychology*, 17, 349–362.
- Searle, J. R. (1975). The logical status of fictional discourse. *New Literary History*, 6, 319–332.
- Springer, K., & Murphy, G. L. (1992). Feature availability in conceptual combination. *Psychological Science*, 3, 111–117.
- Stewart, A. J., Haigh, M., & Kidd, E. (2009). An investigation into the online processing of counterfactual and indicative conditionals. *Quarterly Journal of Experimental Psychology*, 62, 2113–2125.
- Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence processing: Intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research*, 29, 581–595.
- Traxler, M., McElree, B., Williams, R. S., & Pickering, M. J. (2005). Context effects in coercion: Evidence from eye-movements. *Journal of Memory and Language*, 53, 1–25.
- Turner, M., & Fauconnier, G. (1998). Conceptual integration in counterfactuals. In J.-P. Koenig (Ed.), *Conceptual Structure, Discourse, and Language, II*. Stanford: Center for the Study of Language and Information.
- Urbach, T. P., & Kutas, M. (2010). Quantifiers more or less quantify online: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, 63, 158–179.
- Urrutia, M., de Vega, M., & Bastiaansen, M. C. M. (2012). Understanding counterfactuals in discourse modulates ERP and oscillatory gamma rhythms in the EEG. *Brain Research*, 1455, 40–55.
- Van Berkum, J. J. A. (2010). The brain is a prediction machine that cares about good and bad – Any implications for neuropragmatics? *Italian Journal of Linguistics*, 22, 181–208.
- 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. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443–467.
- Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11, 657–671.
- Van Berkum, J. J. A. (2009). The neuropragmatics of “simple” utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke: Palgrave Macmillan.
- Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, 17, 701–718.
- Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 394–417.
- Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, 97, 279–293.
- Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye movements when reading about plausible and impossible events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34, 1001–1010.
- Wittgenstein, L. (1958). *Philosophische Untersuchungen, Philosophical Investigations* (2nd ed.). Oxford: Basil Blackwell.
- Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 920–933.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162–185.