



# Marking the counterfactual: ERP evidence for pragmatic processing of German subjunctives

Eugenia Kulakova\*, Dominik Freunberger and Dietmar Roehm

Centre for Cognitive Research, University of Salzburg, Salzburg, Austria

## Edited by:

Edmund C. Lalor, Trinity College  
Dublin, Ireland

## Reviewed by:

Jonathan Fugelsang, University of  
Waterloo, Canada

Alan James Power, University of  
Cambridge, UK

## \*Correspondence:

Eugenia Kulakova, Centre for  
Cognitive Neuroscience, University  
of Salzburg, Hellbrunnerstrasse 34,  
5020 Salzburg, Austria  
e-mail: [eugenia.kulakova@sbg.ac.at](mailto:eugenia.kulakova@sbg.ac.at)

Counterfactual conditionals are frequently used in language to express potentially valid reasoning from factually false suppositions. Counterfactuals provide two pieces of information: their literal meaning expresses a suppositional dependency between an antecedent (*If the dice had been rigged...*) and a consequent (*...then the game would have been unfair*). Their second, backgrounded meaning refers to the opposite state of affairs and suggests that, in fact, the dice were not rigged and the game was fair. Counterfactual antecedents are particularly intriguing because they set up a counterfactual world which is known to be false, but which is nevertheless kept to when evaluating the conditional's consequent. In the last years several event-related potential (ERP) studies have targeted the processing of counterfactual consequents, yet counterfactual antecedents have remained unstudied. We present an EEG/ERP investigation which employed German conditionals to compare subjunctive mood (which marks counterfactuality) with indicative mood at the critical point of mood disambiguation via auxiliary introduction in the conditional's antecedent. Conditional sentences were presented visually one word at a time. Participants completed an acceptability judgment and probe detection task which was not related to the critical manipulation of linguistic mood. ERPs at the point of mood disambiguation in the antecedent were compared between indicative and subjunctive. Our main finding is a transient negative deflection in frontal regions for subjunctive compared to indicative mood in a time-window of 450–600 ms. We discuss this novel finding in respect to working memory requirements for rule application and increased referential processing demands for the representation of counterfactuals' dual meaning. Our result suggests that the counterfactually implied dual meaning is processed without any delay at the earliest point where counterfactuality is marked by subjunctive mood.

**Keywords:** conditionals, counterfactuals, EEG/ERP, linguistic mood, pragmatics, subjunctive

## INTRODUCTION

Counterfactual conditionals are frequently used in everyday language to convey potential dependencies between antecedents and consequents which have never been realized. A counterfactual conditional "*If Horst had not bought a pony, he would be miserable*" expresses a positive relation between Horst (suppositionally) not buying a pony and his (suppositional) misery. However, most people are able to infer from this sentence that Horst did in fact buy a pony (and is not miserable at all), even in absence of prior information about Horst (Thompson and Byrne, 2002). This notion of a dual meaning (suppositional and factual) has made the topic of counterfactual conditionals interesting for philosophers, linguists and psychologists alike and accounts for increasing attention of neurocognitive and psycholinguistic research on counterfactual sentence processing (Ferguson et al., 2008; Van Hoeck et al., 2010, 2012, 2013; Nieuwland and Martin, 2012; Nieuwland, 2012a,b; Urrutia et al., 2012; Kulakova et al., 2013).

In this recent line of neuroscientifically informed language research, counterfactual conditionals have been characterized by past tense and subjunctive mood. It has been argued before that it is primarily the pragmatics of subjunctive mood that marks the antecedent (i.e., the *if*-part) of the counterfactual conditional as factually false (Stalnaker, 1975). The subjunctive mood signals an inconsistency between the antecedent and the interlocutors' factual world knowledge. In contrast, the use of indicative mood ("*If Helga got a cat, then she is happy now.*") in conditionals suggests compatibility between world knowledge and the proposition expressed by the antecedent. Helga might have bought a cat, or not, we don't know. Thus, even if the hearer does not already know the conflicting facts targeted by the counterfactual antecedent, she can infer them from the speaker's use of subjunctive mood. The factual information conveyed by counterfactuals (i.e., counterfactual antecedent falsity) has initially been identified as a presupposition phenomenon (Levinson, 1983; Fauconnier, 1994), although alternative accounts in terms of presuppositional and scalar implicatures

have been presented recently (Verstraete, 2008; Leahy, 2011). All accounts agree that a pragmatic inference process is initiated to arrive at the (backgrounded) factual meaning of a counterfactual antecedent.

In a cognitive semantic form, the dual representation of counterfactuals was described within the mental spaces framework of Fauconnier (1994). Fauconnier introduces the concept of mental spaces, domains which are used to represent our mental (or subjectively accessible) representation of linguistic content. Counterfactual conditionals are considered as space builders, introducing new spaces as they go. Crucially, the parent space (usually representing the real world) and the counterfactually triggered mental space must contain incompatible information. The mental spaces account of counterfactuals captures the intuitive subjective experience during counterfactual sentence processing. However, it is difficult to approach the implications of this framework in an experimental manner. For example, the theory does not indicate whether a counterfactual is more difficult or costly to process than a sentence which does not open a new mental space. Furthermore, it does not indicate at which point of incremental sentence processing the counterfactual mental space is built up and for how long it is up-held.

Studies of online sentence processing have shown that subjects rapidly represent the factual state of affairs implied by counterfactual conditionals. Subjunctive but not indicative conditionals (*If A then B*) were shown to prime the implied factual event ( $\neg A$  and  $\neg B$ ), resulting in reading time benefits for its subsequent presentation (Santamaría et al., 2005). An eye-tracking study presented either counterfactual conditionals (*If the racing driver had won the race, ...*) or factual sentences (*Because the racing driver had won the race, ...*) as a context manipulation followed by either consistent or inconsistent declarative sentences (Ferguson, 2012). Importantly, the counterfactual continuation (*In the evening the racings driver's team celebrated<sub>inconsistent</sub>/commiserated<sub>consistent</sub>*) was either consistent or inconsistent with the implied factual information of the preceding counterfactual. The fact that subjects detected the inconsistency in the counterfactual condition (as indexed by more first-pass regressions out and back into the critical regions and longer total reading times of the pre-critical region) shows that subjects had access to the implied factual meaning of the counterfactual.

During the recent years an increasing amount of Event-Related Potential (ERP) studies has targeted counterfactual sentence processing. The neurocognitive approach to counterfactuals has primarily focused on the semantic N400 effect of world-knowledge violation in factual compared to counterfactual contexts (Ferguson et al., 2008; Nieuwland and Martin, 2012; Nieuwland, 2012a,b). Nieuwland and Martin (2012) demonstrated that the processing effort of a counterfactual consequent (as indexed by the N400 amplitude elicited by a critical word) is modulated by its consistency with the conditionals' antecedent rather than general world knowledge. This finding was furthermore demonstrated to hold for absolutely absurd consequences (*If dogs had gills, Dobermans would breathe under water*). The ERP elicited at the critical word was comparable to a factually true sentence (*Because fishes have gills, tuna breathe under water*). However, a pronounced N400 effect emerged at the critical

word when the licensing counterfactual antecedent was omitted (Nieuwland, 2012b). This suggests that the (factually false) propositions expressed by counterfactual antecedents are quickly incorporated into the discourse, licensing subsequent conclusions which are factually false but consistent with a locally established context.

These results taken together suggest that both the factual and suppositional meaning of a counterfactual are accessed rapidly, as expected from a double representation account. If this is true, it should be possible to observe effects of increased processing costs arising from the dual counterfactual representation at the point of linguistic mood disambiguation, since counterfactuality is for the first time marked by subjunctive mood use in the conditionals' antecedent. However, to our knowledge no study has yet investigated the immediate effect of counterfactuality introduction in the antecedent of a counterfactual conditional. The reason why research has neglected counterfactual antecedents processing for a long time may lie in the fact that until now ERP studies of counterfactuals have relied on factual "Because ..." constructions as the contrast conditions for counterfactuals, which makes it difficult to interpret any effect at the auxiliary because up to the critical point sentences already differ in length and wording. A recent proposal from imaging research on counterfactual sentence processing suggests that the hypothetical conditional as marked by indicative mood provides a more appropriate control condition, paralleling the counterfactual in the suppositional ontological status as well as the If-then structure (Kulakova et al., 2013). The only remaining difference between these conditions is linguistic mood, introducing the counterfactuals' antagonism to factual events (see also Stewart et al., 2009).

In the present study we investigated the neural consequences of counterfactuality introduction in the antecedent of German conditionals. We compared the ERPs elicited by auxiliary verbs (i.e., grammatical entities that carry information about linguistic mood) marking subjunctive and indicative mood. To our knowledge this was the first attempt to experimentally pinpoint the process of counterfactuality marking in a subjunctive antecedent. We used an electroencephalographic (EEG) measure which is very sensitive to fast, automatic, and unconscious processes during sentence processing and provides detailed information about the time-course of neuronal events. We expected to observe an ERP difference reflecting the process of counterfactual implication detection leading to a dual representation in subjunctive compared to indicative mood.

Opposite to the English language, German allows the implementation of this comparison in a very straightforward way. Subjunctive mood in past tense has a distinct morphology that, contrary to English, does not comprise a double layer of past tense which inevitably results in sentence length differences. Another advantage of German syntax that facilitates the investigation of mood-disambiguation at this early point comes from the fact that the mood-disambiguating auxiliary takes the final position in the antecedent where propositional content is already conveyed. See an example of our stimulus material in **Table 1**. Both conditions are identical in the first four words which set up the suppositional proposition of rigged dice (*Wenn die Würfel gezinkt ...*). The next word is the auxiliary verb which introduces linguistic mood:

**Table 1 | Example of German stimulus sentences (critical word underlined) with literal and analogous English translations.****Subjunctive (counterfactual)**

Wenn die Würfel gezinkt wären, dann wäre das Spiel fair/unfair.  
*If the dice rigged were<sub>subjunctive</sub>, then would be the game fair/unfair.  
 "If the dice had been rigged, then the game would have been fair/unfair."*

**Indicative**

Wenn die Würfel gezinkt waren, dann war das Spiel fair/unfair.  
*If the dice rigged were<sub>indicative</sub>, then was the game fair/unfair.  
 "If the dice were rigged, then the game was fair/unfair."*

In the counterfactual condition "wären" marks subjunctive mood and allows the reader to infer that although it is supposed that the dice were rigged (suppositional meaning), they in fact were not (factual meaning). In the indicative condition "waren" introduces the suppositional reading only, allowing no inferences about the factual state of affairs.

We incorporated our experimental manipulation of linguistic mood in an acceptability judgment and probe detection task, so that consciously inferring the counterfactually implied factual information was not required for correct task performance. This was intended in order to investigate the automatic (i.e., task-irrelevant) effect of counterfactual sentence processing occurring at the point of mood disambiguation via the introduction of the auxiliary in the conditionals' antecedent. We used German conditional sentences describing causal and definitional regularities which have been demonstrated to trigger a counterfactual reading when presented in subjunctive mood (Thompson and Byrne, 2002). We measured ERPs elicited at the critical antecedent-final auxiliary but also report the effects of the sentence-final manipulation of truth. Since the targeted process of counterfactuality processing was semantic/pragmatic in nature we did not expect to observe the critical effect to occur early, but rather in time-window where effects of semantic manipulations (i.e., N400, LAN, P600) are usually reported. Potential effects of visual dissimilarity of the compared auxiliaries should, if occurring at all, manifest in rather early differences starting at around 150 ms post onset (Sauseng et al., 2004).

**MATERIALS AND METHODS****PARTICIPANTS**

Sixteen German native speakers (7 female; mean age 23.5; age range 18–36 years) participated in the ERP study. All subjects were right-handed according to an adopted German version of the Edinburgh Inventory (Oldfield, 1971) and had normal or corrected-to-normal vision. Subjects gave informed consent before the experiment and were paid 20€ for participation. All methods conformed to the Code of Ethics of the World Medical Association (Declaration of Helsinki) (see Mayrhauser et al., 2014).

**MATERIALS**

Stimulus material consisted of 52 sets of German conditionals describing social, physical or biological regularities. The experimental factor of interest was MOOD (two levels: indicative, subjunctive). Furthermore, the sentence-final content word

varied as a function of TRUTH (two levels: true, false) in order to balance acceptability. Two auxiliary verbs forms, *haben* (to have) and *sein* (to be), were used in equal amounts to make the stimuli less monotonous. Their past tense plural forms are visually maximally similar between past indicative and subjunctive (*hatten/hätten, waren/wären*), only differing in the diacritic umlaut on the first vowel. Since this could have added an effect of auxiliary due to possible differences in naturalness, cloze or word frequency effects, we carried out additional tests to specify these parameters and also included auxiliary type (factor AUX, two levels: to be, to have) in further analyses. Each of the 52 thematic vignettes was presented twice per subject in different counter-balanced conditions, resulting in 104 items in total. Additionally, the experiment contained 208 declarative (non-conditional) filler sentences which were also balanced in acceptability (e.g., *No vegetarian likes eating roast/vegetables*).

**Naturalness rating**

To investigate possible effects of perceived naturalness, we conducted a rating-task of our stimulus material in respect to the four auxiliaries employed. In an internet-based survey 50 participants (none of which took part in the EEG-experiment) rated the auxiliaries of the antecedents (consequents were not presented to avoid confounding effects of plausibility) for their naturalness on a 7 point scale. A  $2 \times 2$  repeated measures ANOVA with the factors MOOD and AUX revealed a main effect of MOOD [ $F_{(1, 48)} = 163.2, p < 0.001$ ] and a MOOD  $\times$  AUX interaction [ $F_{(1, 48)} = 14.7, p < 0.001$ ]. Resolving the interaction by AUX indicated that naturalness ratings were higher for subjunctive both for "to be" [ $T_{(48)} = 4.6, p < 0.001$ ] and "to have" [ $T_{(48)} = 12.8, p < 0.001$ ]. See Table 2 for detailed behavioral results.

**Completion task**

We also tested predictability differences between indicative and subjunctive auxiliaries employed in our task. In an internet-based survey a cohort of 58 subjects (none of which took part in the EEG-experiment or carried out the naturalness rating) completed 30 randomly selected sentences from our stimulus material truncated before the critical auxiliary. The subjunctive past tense continuation was chosen with a mean probability of 0.13, the indicative past auxiliary was almost never chosen ( $< 0.01$ ). Instead, subjects chose the more colloquial future or present indicative forms of the respective auxiliary, but no other word forms. See Table 2 for detailed cloze values. Table 2 also presents the frequencies (per million) of the employed auxiliaries derived from the SUBTLEX-DE database (Brysbaert et al., 2011).

**PROCEDURE**

Trials began with a fixation cross presented for 1500 ms. After 900 ms blank screen, sentences were presented word-by-word with a single word duration of 400 ms (nominal phrases 500 ms) and an inter-stimulus interval of 200 ms. The critical words were presented with a comma, the last word of each sentence was presented with a period. Each sentence was followed by an acceptability judgment where a green question mark prompted subjects to give a yes/no acceptability judgment. This was followed by a probe detection task in which subjects judged if a

**Table 2 | Stimulus properties and behavioral results split by mood (indicative/subjunctive) and auxiliary (to be/to have).**

			Indicative		Subjunctive	
			To be	To have	To be	To have
Frequency			711	363	166	242
Naturalness			4.1 (1.7)	3.5 (1.2)	5.3 (1.4)	5.7 (0.8)
Cloze			0	0	0.08	0.18
Acceptability judgment	True	% accepted	95.1 (7)	95.1 (7)	94.7 (8)	93.1 (7)
		RT in ms	577 (245)	712 (260)	620 (286)	712 (245)
	False	% accepted	4.4 (7)	3.5 (7)	5.7 (8)	4.2 (11)
		RT in ms	606 (289)	742 (308)	730 (326)	724 (341)
Probe detection	True	% detected	96.9 (5)	99.6 (2)	98.7 (3)	99.0 (4)
		RT in ms	884 (243)	853 (254)	852 (233)	946 (290)
	False	% detected	99.5 (2)	99.1 (2)	98.4 (3)	99.5 (2)
		RT in ms	922 (314)	994 (234)	1026 (300)	945 (248)

Frequencies per million, offline ratings of naturalness ( $N = 50$ ) and cloze probability ( $N = 58$ ). Online acceptability and probe detection responses ( $N = 16$ ) with corresponding reaction times (RT) for true and false sentences. Standard deviations are given in parentheses.

single displayed word had been presented in the previous sentence. Maximal response duration for each task was 2000 ms. Participants responded by pressing the left or right shift-key on a computer keyboard. The assignment of response buttons was counterbalanced across participants. Inter-trial interval was 800 ms. Experimental and filler sentences (312 trials in total) were pseudo-randomized and presented in 6 blocks with 52 different items per block. Half of the participants read the sentences in reverse order to control for sequence effects.

### EEG RECORDING AND PREPROCESSING

EEG was recorded from 25 Ag/AgCl electrodes mounted in an elastic cap (Easy Cap International, Herrsching-Breitbrunn, Germany) according to the 10/20 system (Jasper, 1958). Electrodes included: AFz, Fz, FCz, Cz, CPz, Pz, POz, F7/8, F3/4, FC5/6, FC1/2, CP5/6, CP1/2, P7/8, P3/4, O1/2. The EEG-signal was sampled at 500 Hz with a low-pass filter at 250 Hz and a software notch filter (50 Hz). Data were recorded with respect to the left mastoid reference and electrode AFz served as ground electrode. The horizontal electro-oculogram (HEOG) was recorded from electrodes at the outer canthus of each eye. The vertical electro-oculogram (VEOG) was recorded from electrodes placed above and beneath the left eye. Scalp impedances were kept below 5 k $\Omega$ .

Offline, all electrodes were re-referenced to the average activity of the left and right mastoids, before an independent component analysis (ICA) was applied to correct ocular artifacts. After correction, remaining artifacts (EOG, movement, technical) were marked manually. In total, 5.1% of indicative 7.7% of subjunctive trials had to be rejected. The signal was band-pass filtered from 0.3–20 Hz in order to exclude slow drifts and then was segmented into epochs from –300 to 800 ms around the critical words. The signal was averaged for each condition and each participant before grand averages were computed across all participants.

For the statistical analysis of acceptability and probe detection, mean acceptability (% accepted) and probe detection accuracy (% detected) and the corresponding reaction times

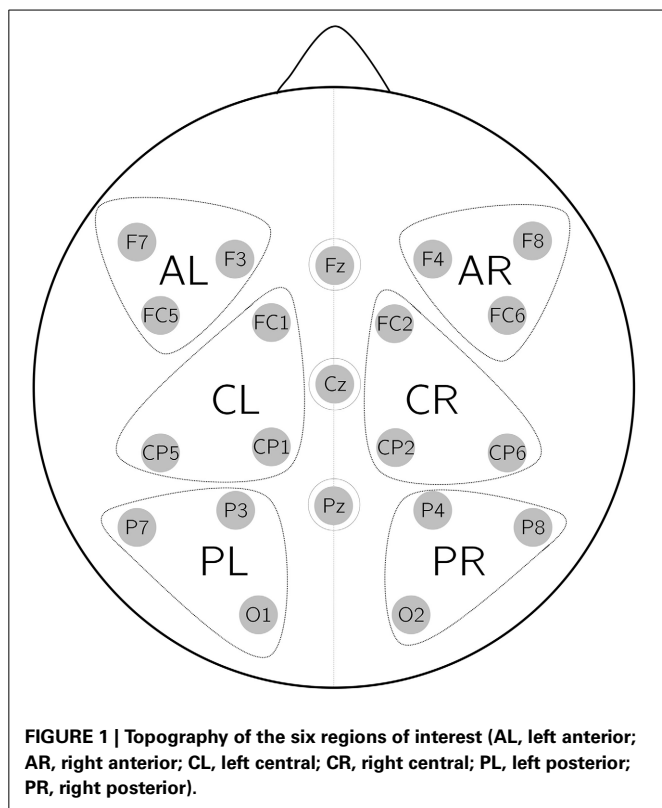
were subjected to repeated measure ANOVAs with the factors MOOD (indicative/subjunctive), TRUTH (true/false) and AUX (to have/to be).

Based on visual inspection statistical analyses of the ERPs were carried out in two time windows: 200–400 ms and 450–600 ms. For both time windows the mean amplitude was computed per subject, condition, and electrode. Mean amplitude values were submitted to a  $2 \times 2 \times 6$  repeated measures ANOVA with the factors MOOD, AUX and ROI (regions of interest). The factor TRUTH was omitted because it played no role at the critical position. Six ROIs were calculated from the mean value of three electrodes each: anterior left (AL; F3, F7, FC5), anterior right (AR; F4, F8, FC6), central left (CL; FC1, CP1, CP5), central right (CR; FC2, CP2, CP6), posterior left (PL; P3, P7, O1), and posterior right (PR; P4, P8, O2), see **Figure 1** for the topography of the ROIs. Another  $2 \times 2 \times 3$  ANOVA was performed with the factor MID (midline electrodes) where three midline electrodes were treated as separate ROIs (Fz, Cz, Pz). *P*-values with more than one degree of freedom in the numerator were corrected according to Greenhouse-Geisser. Only significant interactions were resolved. To disentangle the effects emerging in the statistics pairwise comparisons were computed and are reported at an FDR-corrected threshold (Benjamini and Hochberg, 1995). Effects of AUX, ROI, and MID are presented but not discussed unless there is an interaction with MOOD.

Given the employed target manipulation (Steinhauer and Drury, 2012) of the critical auxiliary, no systematic differences between the baseline intervals were expected. Instead of using a baseline correction which could carry over transient signal differences from the baseline interval to the critical time window, we employed a filter which serves to remove unsystematic pre-stimulus differences due to slow drifts (below 0.3 Hz). The baseline interval was checked for significant differences in steps of 50 ms from –200 to 100 ms around the onset of the critical word to demonstrate the absence of baseline differences.

As a further check of data quality the sentence-final words were analyzed in the time windows of 300–500 ms and 500–800 ms





with the expectation to observe an N400/P600 effect commonly known to reflect semantic violations. All additional analyses are reported in the Supplementary Material.

## RESULTS

### BEHAVIORAL RESULTS

#### Acceptability judgment

Subjects rated true sentences as acceptable and false sentences as unacceptable. The 2 (MOOD)  $\times$  2 (TRUTH)  $\times$  2 (AUX) repeated measures ANOVA revealed only a main effect of TRUTH [ $F_{(1, 15)} = 2005.6, p < 0.001$ ] with higher acceptability rates for true compared to false sentences. The comparison of reaction times for acceptability ratings showed a main effect of AUX [ $F_{(1, 15)} = 26.3, p < 0.001$ ] and a significant AUX  $\times$  MOOD interaction [ $F_{(1, 15)} = 4.8, p < 0.05$ ]. Resolving the interaction by AUX indicated that only in the auxiliaries of “to be” subjunctive mood led to significantly increased reaction times [ $T_{(15)} = 2.5, p < 0.05$ ]. A possible reason for this effect is a larger length difference between indicative and subjunctive sentence-final words after “to be” (difference of 0.24 characters) than after “to have” (difference of 0.05 characters).

#### Probe detection

Performance in the probe detection task was almost perfect. A 2  $\times$  2  $\times$  2 repeated measures ANOVA showed neither main effects nor interactions of the explanatory variables. The analysis of probe detection speed showed a main effect of TRUTH [ $F_{(1, 15)} = 12.2, p < 0.01$ ] and an AUX  $\times$  MOOD  $\times$  TRUTH interaction [ $F_{(1, 15)} = 6.7, p < 0.05$ ]. Resolving the interaction by

AUX revealed a significant MOOD  $\times$  TRUTH interaction for “to have” [ $F_{(1, 15)} = 8.8, p < 0.01$ ] but not for “to be” [ $F_{(1, 15)} = 2.8, p = 0.11$ ]. Further paired comparisons showed that reaction times in true “to have” sentences were faster only in indicative [ $T_{(15)} = 3.9, p < 0.01$ ] but not subjunctive mood [ $T_{(15)} < 1$ ]. Detailed behavioral responses with corresponding reaction times can be found in **Table 2**.

### ERP RESULTS

Detailed test-statistics are shown in **Table 3**. The 2 (MOOD)  $\times$  2 (AUX)  $\times$  6 (ROI) repeated measure ANOVA for the lateral electrodes in the 200–400 ms time window revealed that subjunctive mood elicited a significantly more positive deflection compared to indicative as indicated by a main effect of MOOD. Midline analyses in the same time window showed significant main effects of MOOD and MID and a significant AUX  $\times$  MID interaction. However, resolving this interaction did not reveal any electrodes at which the AUX effect reached significance [ $Fz: T_{(15)} = 1.34, p_{FDR} = 0.60; Cz/Pz: T_{(15)} < 1$ ].

In the 450–600 ms time window the analysis of lateral ROIs showed main effects of MOOD and ROI and significant MOOD  $\times$  ROI and AUX  $\times$  ROI interactions. Former indicated that the stronger negativity elicited by subjunctive mood was most pronounced at left anterior and central electrodes and absent in right and posterior regions. Resolving the AUX  $\times$  ROI interaction did not reveal any ROI which showed a significant effect auf AUX [AL/CL/CR/PR:  $T_{(15)} < 1$ ; AR:  $T_{(15)} = 1.31, p_{FDR} = 0.67$ ; PL:  $T_{(15)} = 1.27, p_{FDR} = 0.67$ ]. The midline ANOVA showed a main effect of MOOD and MID with a significant MOOD  $\times$  MID and AUX  $\times$  MID interaction. The former indicated a frontal distribution of the mood effect (i.e., stronger negativity for subjunctive compared to indicative). Resolving the latter did not disclose any electrodes at which the effect of AUX reached significance [ $Fz: T_{(15)} = 1.12, p_{FDR} = 0.82; Cz/Pz: T_{(15)} < 1$ ]. **Figure 2** shows the grand average ERP waveforms at selected electrode sites for both levels of the experimental condition MOOD, time locked to the critical point of mood disambiguation via auxiliary introduction.

The analysis of the sentence-final content word showed a clear N400/P600 pattern which was expected given our manipulation of truth and the acceptability rating task. None of the pre-critical time-windows showed effects of MOOD or AUX. Detailed test-statistics of the additional analyses can be found in the Supplementary Material.

### DISCUSSION

The present ERP experiment examined conditional sentence processing at the critical point of linguistic mood introduction in the antecedent of conditionals. We expected to observe an ERP effect reflecting the process of counterfactuality marking in subjunctive compared to indicative antecedents. This implication detection process is required to establish a dual representation of counterfactual meaning and was expected to be automatic and thus to occur in a task which did not involve explicit factual implication computation. In accordance with our expectations we found a positive deflection for subjunctive compared to indicative auxiliaries in the time-window between 200–400 ms, followed by a left frontal negativity in a time window of 450–600 ms.

**Table 3 | Test-statistics of the ERPs elicited by the critical auxiliaries in two time windows.**

ANOVA lateral electrodes	Time window							
	200–400 ms				450–600 ms			
	<i>df</i>	<i>F</i>	<i>p</i>	<i>sig</i>	<i>df</i>	<i>F</i>	<i>p</i>	<i>sig</i>
MOOD	1, 15	26.34	<0.001	***	1, 15	4.36	<0.05	*
AUX	1, 15	<1	0.97		1, 15	<1	0.96	
ROI	5, 75	15.27	<0.001	***	5, 75	5.58	<0.01	**
AUX × MOOD	1, 15	<1	0.46		1, 15	1.33	0.27	
MOOD × ROI	5, 75	2.46	0.08		5, 75	3.07	<0.05	*
AUX × ROI	5, 75	2.53	0.09		5, 75	3.06	<0.05	*
AUX × MOOD × ROI	5, 75	<1	0.55		5, 75	1.70	0.18	
<b>Resolution by ROI</b>					<b><i>df</i></b>	<b><i>T</i></b>	<b><i>p</i><sub>FDR</sub></b>	<b><i>sig</i><sub>FDR</sub></b>
MOOD (AL)					15	3.07	<0.05	*
MOOD (AR)					15	1.89	0.12	
MOOD (CL)					15	2.23	0.12	
MOOD (CR)					15	1.86	0.12	
MOOD (PL)					15	1.36	0.23	
MOOD (PR)					15	<1	0.59	
<b>ANOVA midline electrodes</b>	<b><i>df</i></b>	<b><i>F</i></b>	<b><i>p</i></b>	<b><i>sig</i></b>	<b><i>df</i></b>	<b><i>F</i></b>	<b><i>p</i></b>	<b><i>sig</i></b>
MOOD	1, 15	15.42	<0.001	***	1, 15	4.70	<0.05	*
AUX	1, 15	<1	0.57		1, 15	<1	0.78	
MID	2, 30	19.18	<0.001	***	2, 30	13.71	<0.001	***
AUX × MOOD	1, 15	<1	0.97		1, 15	1.34	0.27	
MOOD × MID	2, 30	1.81	0.19		2, 30	3.90	<0.05	*
AUX × MID	2, 30	3.89	<0.05	*	2, 30	4.24	<0.05	*
AUX × MOOD × MID	2, 30	<1	0.55		2, 30	<1	0.66	
<b>Resolution by MID</b>					<b><i>df</i></b>	<b><i>T</i></b>	<b><i>p</i><sub>FDR</sub></b>	<b><i>sig</i><sub>FDR</sub></b>
MOOD (Fz)					15	3.58	<0.01	**
MOOD (Cz)					15	1.99	0.10	
MOOD (Pz)					15	<1	0.36	

Main effects and interactions of the performed ANOVAs and resolution of significant interactions with Mood by region of interest (ROI) and midline (MID) electrodes, respectively. AUX, auxiliary; AL, anterior left; AR, anterior right; CL, central left; CR, central right; PL, posterior left; PR, posterior right.

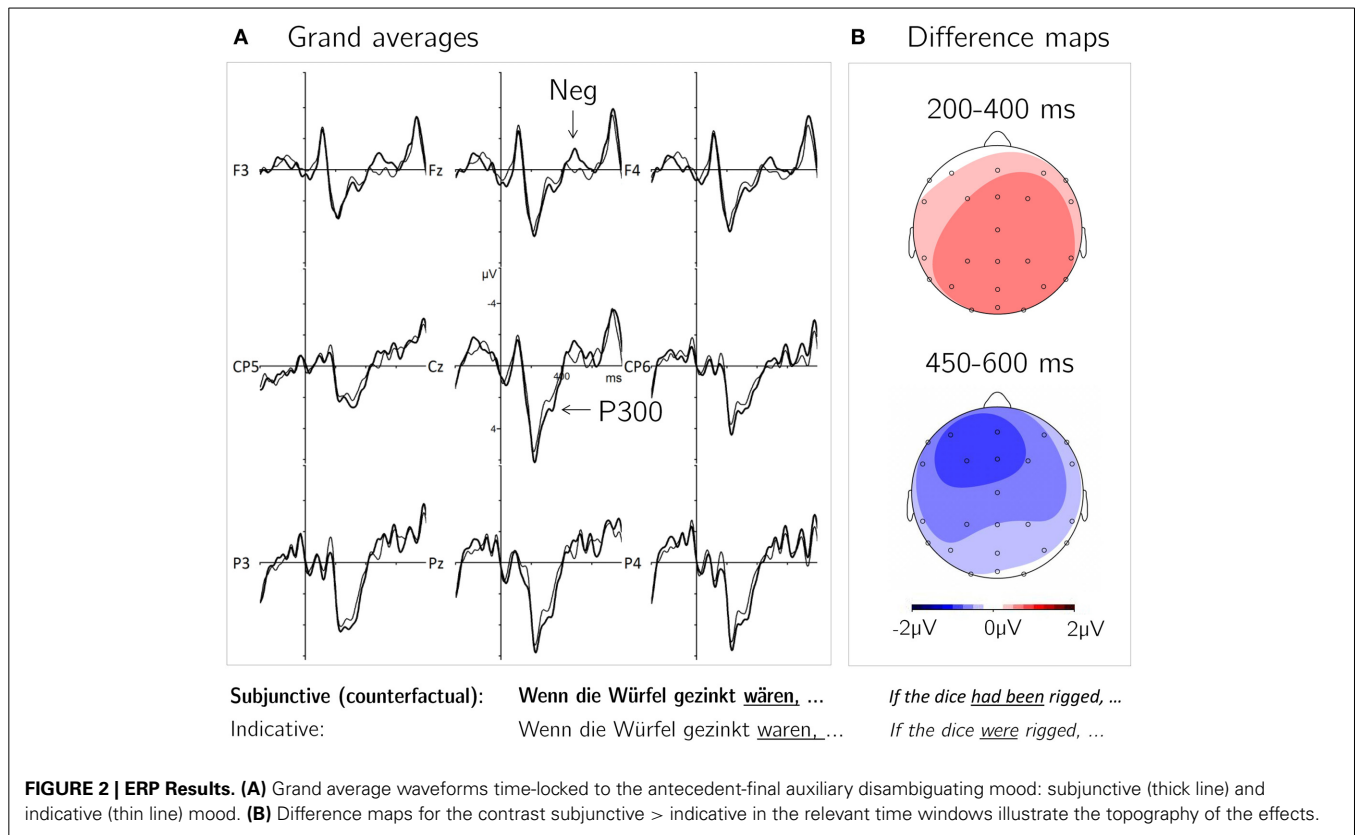
### PREDICTION-RELATED P300

We interpret the early centro-parietal positivity for subjunctive relative to indicative auxiliaries as an instance of the prediction-related P300, since it has a similar time-course and distribution as described in previous studies (Roehm et al., 2007; Molinaro and Carreiras, 2010; Vespignani et al., 2010). This ERP component has been shown to reflect the integration of an expected lexical element into the prior context. When there is a successful match of the predicted and the actual input, a P300 is generated. Cloze values and naturalness-ratings support the conjecture: in our stimulus set the subjunctive continuation was anticipated with a greater likelihood and rated to sound more natural than the indicative auxiliary. Thus, the antecedent's context triggered a significantly stronger expectation of the subjunctive auxiliaries, the integration of which elicited a P300 effect in comparison to the less expected indicative auxiliary. This effect arises from the

cloze values of the employed stimuli but should not be replicated if cloze values between indicative and subjunctive auxiliaries were matched.

### LEFT ANTERIOR NEGATIVITY AND DEFINITENESS

Interestingly, the negative frontal effect for subjunctive vs. indicative auxiliaries shows that, although locally predicted, counterfactuals pose increased processing demands. Since lexical prediction presumably does not go beyond semantics, the pragmatic implication carried by the subjunctive word forms still required additional processing. The observed transient negativity from 450–600 ms for subjunctive auxiliaries bears topographical and temporal resemblance to the left anterior negativity (LAN) effect as often observed for function words as employed in our study. Generally, the LAN has been proposed to index increased syntactic working memory demands imposed



by un-canonical sentence structure (Rösler et al., 1998; Matzke et al., 2002). Recently, a transient LAN effect was observed for regular compared to irregular German past participles and was thus proposed as a marker of syntactic rule application (Krott and Lebib, 2012). In our task, however, the employed correct form auxiliaries did not differ functionally between indicative and subjunctive mood (neither in terms of inflection, conveyed syntactic complexity nor task demands). Thus, the increased working-memory demand indexed by the observed negativity effect is most parsimoniously explained as stemming from the pragmatic process of counterfactual implication detection, mirroring the application of a pragmatic rather than syntactic rule.

A recent finding supports the link between the present results and the topic of implication detection in terms of presupposition processing. German definite articles in noun-phrases have been reported to trigger a significantly stronger negative deflection than indefinite articles in the time-window of 400–700 ms with a left anterior topography very similar to the one presently observed (Schumacher, 2009). This effect has been argued to reflect increased working memory demands for the integration of a definite referent into discourse. Notably, definite articles and counterfactuals have both been proposed to be presupposition triggers (Levinson, 1983), since both elicit a pragmatic inference process in order to make sense of the use of a specific expression given the shared beliefs of speaker and hearer. The presupposition attached to the use of the definite article “the” introduces the assumption that the referent of the noun following “the”

exists and is contextually salient. This parallels the process by which subjunctive mood marks the proposition of the antecedent as false, simultaneously implying its opposite as factually true, thus referring to a specific factual event. Indefinite articles and indicative mood, on the other hand, do not convey any definiteness about the introduced proposition, and thus lack reference to any specific discourse entity. Taken together, these results suggest that the amplitude of the observed negativity is sensitive to the pragmatic process of counterfactuality marking via implication detection.

#### EFFECTS OF FREQUENCY AND CLOZE

Since the frequencies of the employed subjunctive auxiliaries are generally lower than those marking indicative mood, an effect of word frequency could be expected to drive our main finding. However, the LAN reported for closed-class words has been shown to be frequency independent (Brown et al., 1999; Molinaro et al., 2008). Only one study reported a more pronounced negativity for very highly frequent compared to high, medium and low frequency words (Münte et al., 2001). Thus, our results show if any, then the opposite pattern of a frequency effect. There is therefore no indication that the negativity effect reported in our data stems from the frequencies of the employed auxiliaries.

A similar line of argumentation applies to a possible confounding effect of cloze value and rated naturalness. Less expected words are usually associated with a stronger negativity compared to highly anticipated continuations of a sentence in terms of an N400 effect (Federmeier et al., 2007). Since in our stimuli the

indicative auxiliary received lower cloze values and naturalness ratings our main finding of a stronger negativity for subjunctive auxiliaries is contrary to an expected effect driven by cloze. Thus, the confounding with cloze/naturalness and frequency in our material rather leads to an underestimation of the observed effect associated with processing counterfactuals than being an alternative explanation for it. Since the N400 is known to be most pronounced over central and posterior electrode sites (e.g., Kutas and Federmeier, 2011) such an overlap would be strongest at these regions and thus could be shifting the observed negativity effect of counterfactuality toward anterior sites. In this case further studies with cloze-values matched for mood might find stronger negativity effects of counterfactuality with broader distributions than presently observed.

### REFERENTIALY INDUCED FRONTAL NEGATIVITY

Another stance toward the dual meaning of counterfactuals is referential plurality. Two meanings, one literal suppositional and the other backgrounded factual, are conveyed at the point where subjunctive mood suggests a counterfactual reading of the antecedent's proposition. Indicative antecedents, on the other hand, carry only their literal suppositional meaning. This links the present findings to literature on referential ambiguity where processing pronouns with more than one definite referent in the salient discourse space has been shown to elicit sustained frontal negative shifts emerging at around 300 ms after stimulus onset (Van Berkum et al., 2007). There are three interpretations of this so called Nref-effect. The negativity could reflect merely noticing the ambiguity, the attempt to resolve it in terms of inference processes, or the attempt to simultaneously keep two competing referential interpretations in working memory, an interpretation which relates the Nref to the sustained LAN effect reflecting increased working memory demands. Since counterfactuals' referential plurality probably does not introduce a sustained ambiguity which has to be resolved at some later point, our finding of a transient negativity for subjunctive auxiliaries might be related to the Nref but rather represent the identification and computation of referential plurality of counterfactuals in terms of their dual meaning.

Taken together, we propose that the increased processing demand for subjunctive mood as indexed by the observed negativity effect reflects the process of counterfactuality marking in terms of implication detection which is relevant for a representation of the counterfactuals' dual meaning. This process appears on an implicit, task-independent level of sentence processing.

### N400/P600 EFFECTS OF SEMANTIC ACCEPTABILITY

Although not in the focus of our investigation, we additionally analyzed the ERPs elicited by the sentence-final content word which established the manipulation of truth in our task. Behavioral results indicate that the acceptability ratings were given in accordance to our truth manipulation. Similarly, the ERPs results show a clear N400/P600 signature for content-words which rendered the sentence false in comparison to those which rendered it true. "False" words were semantically incongruent with preceding context and therefore no contextual facilitation of the lexical entry could be initialized. The N400 reflects

the increased costs of the lexico-semantic access of incongruent compared to congruent sentence-final words (Kutas and Federmeier, 2011). Furthermore, it has been shown that acceptability rating tasks make it more likely for semantic violations to elicit a P600 effect following the N400 (Kuperberg, 2007), which is what we observe in our data. Our N400 results are consistent with previous findings of Nieuwland (2012b) and Nieuwland and Martin (2012) which indicate that the neurocognitive effect of establishing truth-value in counterfactuals depends on the consistency with the preceding antecedent rather than on general world knowledge. Consistently, we did not observe any modulation of the N400 effect by linguistic mood as indicated by the absence of an interaction between these factors.

### CONCLUSION

Our study is to our knowledge the first investigation of counterfactual antecedents using ERPs. We observed that subjunctive auxiliaries which mark counterfactuality in antecedents elicited an early positivity which we interpret as a prediction-related P300 effect, mirroring the increased expectation of a subjunctive compared to an indicative continuation of the antecedents in our stimulus set. More importantly, we observed a negativity effect with a left frontal distribution for subjunctive compared to indicative auxiliaries in the time-window of 450–600 ms. We propose that this effect indicates increased processing demands of counterfactual antecedents triggered by the establishing of the counterfactuals' dual meaning. Our results demonstrate that the dual meaning arises without any delay at the earliest point where counterfactuality is marked.

This result demonstrates the existence of a neural correlate of subjunctive mood processing. This first evidence of a neural marker for counterfactuality will allow further studies to explore the sensitivity to counterfactuality in different contexts and age groups. The negativity effect may for example serve as a developmental precursor for counterfactual sentence comprehension capabilities and should precede children's ability to reason correctly from a counterfactual antecedent without committing the reality bias or resorting to basic conditional reasoning which is characterized by neglecting the factually implied specific circumstances of counterfactuals (Rafetseder and Perner, 2014). Similar applications are conceivable for patients with lesions or clinical populations.

An appealing speculation is that the observed negativity effect is associated with the initiation of inhibitory processes necessary to overcome the reality-bias in counterfactual reasoning. This would lead to the prediction that the strength of the effect depends on contextual salience of factually implied events because more salient stimuli need more inhibition. This motivates subsequent studies which employ absurd counterfactual antecedents as used by Ferguson and Sanford (2008) or Nieuwland (2012b) which explicitly violate factual knowledge (e.g., "If dogs had gills..."), with the expectation to observe even stronger integration costs. Furthermore, correlational investigations of negativity amplitude and duration with individual executive function scores appear promising.

An alternative speculation is that the negativity observed in our study is an index of a pragmatic violation in terms of a process



comparable to presupposition failure. The greatest share of literature reporting a transient LAN (which is, however, followed by a late positivity which is not observed here) comes from ERP studies of morpho-syntactic violations (Kutas and Hillyard, 1983). Although our data do not show a significant late positivity effect and counterfactual antecedents are rated as sounding rather natural, it could be tentatively proposed that the use of subjunctive mood in our materials constitutes a pragmatic violation, since no contradictory factual knowledge is salient to license its use. In that case more salient factual events which are implicitly referred to by the counterfactual antecedent should rather decrease the amplitude of the negativity effect, since they operate as a licenser and facilitate reference assignment. More studies with manipulations of context and task that take into account individual difference in pragmatic sensitivity are required to resolve this interesting issue.

However, all conceivable explanations support an interpretation of the observed negativity effect for subjunctive antecedents in terms of a pragmatic effect related to inference processes associated with counterfactuality marking in order to arrive at the counterfactuals' dual meaning.

## ACKNOWLEDGMENT

The first two authors are financially supported by the Doctoral College "Imaging the Mind" of the Austrian Science Fund (FWF-W1233).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fnhum.2014.00548/abstract>

## REFERENCES

- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300.
- Brown, C. M., Hagoort, P., and ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing: Open and closed-class words. *J. Cogn. Neurosci.* 11, 261–281. doi: 10.1162/089892999563382
- Brybaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., and Böhl, A. (2011). The word frequency effect: a review of recent developments and implications for the choice of frequency estimates in German. *Exp. Psychol.* 58, 412–424. doi: 10.1027/1618-3169/a000123
- Fauconnier, G. (1994). *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511624582
- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., and Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Res.* 1146, 75–84. doi: 10.1016/j.brainres.2006.06.101
- Ferguson, H. J. (2012). Eye movements reveal rapid concurrent access to factual and counterfactual interpretations of the world. *Q. J. Exp. Psychol.* 65, 939–961. doi: 10.1080/17470218.2011.637632
- Ferguson, H. J., and Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: an eye-movement investigation. *J. Mem. Lang.* 58, 609–626. doi: 10.1016/j.jml.2007.06.007
- Ferguson, H. J., Sanford, A. J., and Leuthold, H. (2008). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Res.* 1236, 113–125. doi: 10.1016/j.brainres.2008.07.099
- Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.
- Krott, A., and Lebib, R. (2012). Electrophysiological evidence for a neural substrate of morphological rule application in correct wordforms. *Brain Res.* 1496, 70–83. doi: 10.1016/j.brainres.2012.12.012
- Kulakova, E., Aichhorn, M., Schurz, M., Kronbichler, M., and Perner, J. (2013). Processing counterfactual and hypothetical conditionals: an fMRI investigation. *Neuroimage* 72, 265–271. doi: 10.1016/j.neuroimage.2013.01.060
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49. doi: 10.1016/j.brainres.2006.12.063
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Kutas, M., and Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Mem. Cogn.* 11, 539–550. doi: 10.3758/BF03196991
- Leahy, B. (2011). Presuppositions and antipresuppositions in conditionals. *Proc. SALT* 21, 257–274.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge, UK: Cambridge University Press.
- Matzke, M., Mai, H., Nager, W., Rüsseler, J., and Münte, T. (2002). The costs of freedom: an ERP-study of non-canonical sentences. *Clin. Neurophysiol.* 113, 844–852. doi: 10.1016/S1388-2457(02)00059-7
- Mayrhauser, L., Bergmann, J., Crone, J., and Kronbichler, M. (2014). Neural repetition suppression: evidence for perceptual expectation in object-selective regions. *Front. Hum. Neurosci.* 8:225. doi: 10.3389/fnhum.2014.00225
- Molinaro, N., and Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biol. Psychol.* 83, 176–190. doi: 10.1016/j.biopsycho.2009.12.006
- Molinaro, N., Vespignani, F., Canal, P., Fonda, S., and Cacciari, C. (2008). Cloze probability does not only affect N400 amplitude: the case of complex prepositions. *Psychophysiology* 45, 1008–1012. doi: 10.1111/j.1469-8986.2008.00694.x
- Münte, T. F., Wiering, B. M., Weyerts, H., Szentkuti, A., Matzke, M., and Johannes, S. (2001). Differences in brain potentials to open and closed class words: class and frequency effects. *Neuropsychologia* 39, 91–102. doi: 10.1016/S0028-3932(00)00095-6
- Nieuwland, M. S. (2012a). Establishing propositional truth-value in counterfactual and real-world contexts during sentence comprehension: differential sensitivity of the left and right inferior frontal gyri. *Neuroimage* 59, 3433–3440. doi: 10.1016/j.neuroimage.2011.11.018
- Nieuwland, M. S. (2012b). "If a lion could speak...": online sensitivity to propositional truth-value of unrealistic counterfactual sentences. *J. Mem. Lang.* 68, 54–67. doi: 10.1016/j.jml.2012.08.003
- Nieuwland, M. S., and 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. doi: 10.1016/j.cognition.2011.09.001
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Rafetseder, E., and Perner, J. (2014). Counterfactual reasoning: sharpening conceptual distinctions in developmental studies. *Child Dev. Perspect.* 8, 54–58. doi: 10.1111/cdep.12061
- Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., and Schlewsky, M. (2007). To predict or not to predict: influences of task and strategy on the processing of semantic relations. *J. Cogn. Neurosci.* 19, 1259–1274. doi: 10.1162/jocn.2007.19.8.1259
- Rösler, F., Pechmann, T., Streb, J., Röder, B., and Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *J. Mem. Lang.* 38, 150–176. doi: 10.1006/jmla.1997.2551
- Santamaría, C., Espino, O., and Byrne, R. M. (2005). Counterfactual and semi-factual conditionals prime alternative possibilities. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1149–1154. doi: 10.1037/0278-7393.31.5.1149
- Sauseng, P., Bergmann, J., and Wimmer, H. (2004). When does the brain register deviances from standard word spellings?—An ERP study. *Cogn. Brain Res.* 20, 529–532. doi: 10.1016/j.cogbrainres.2004.04.008
- Schumacher, P. B. (2009). Definiteness marking shows late effects during discourse processing: evidence from ERPs. *Lecture Notes Artif. Intell.* 5847, 91–106. doi: 10.1007/978-3-642-04975-0\_8
- Stalnaker, R. (1975). Indicative conditionals. *Philosophia* 5, 269–286. doi: 10.1007/BF02379021
- Steinhauer, K., and Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain Lang.* 120, 135–162. doi: 10.1016/j.bandl.2011.07.001

- Stewart, A. J., Haigh, M., and Kidd, E. (2009). An investigation into the online processing of counterfactual and indicative conditionals. *Q. J. Exp. Psychol.* 62, 2113–2125. doi: 10.1080/17470210902973106
- Thompson, V. A., and Byrne, R. M. (2002). Reasoning counterfactually: making inferences about things that didn't happen. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 1154–1170. doi: 10.1037/0278-7393.28.6.1154
- Urrutia, M., de Vega, M., and Bastiaansen, M. (2012). Understanding counterfactuals in discourse modulates ERP and oscillatory gamma rhythms in the EEG. *Brain Res.* 1455, 40–55. doi: 10.1016/j.brainres.2012.03.032
- Van Berkum, J. J., Koornneef, A. W., Otten, M., and Nieuwland, M. S. (2007). Establishing reference in language comprehension: an electrophysiological perspective. *Brain Res.* 1146, 158–171. doi: 10.1016/j.brainres.2006.06.091
- Van Hoeck, N., Begtas, E., Steen, J., Kestemont, J., Vandekerckhove, M., and Van Overwalle, F. (2013). False belief and counterfactual reasoning in a social environment. *Neuroimage* 90, 315–325. doi: 10.1016/j.neuroimage.2013.12.043
- Van Hoeck, N., Ma, N., Ampe, L., Baetens, K., Vandekerckhove, M., and Van Overwalle, F. (2012). Counterfactual thinking: an fMRI study on changing the past for a better future. *Soc. Cogn. Affect. Neurosci.* 8, 556–564. doi: 10.1093/scan/nss031
- Van Hoeck, N., Ma, N., Van Overwalle, F., and Vandekerckhove, M. (2010). Counterfactual thinking and the episodic system. *Behav. Neurol.* 23, 225–227. doi: 10.1155/2010/312736
- Verstraete, J.-C. (2008). The nature and origins of counterfactuality in simple clauses: cross-linguistic evidence. *J. Pragmat.* 40, 1865–1895. doi: 10.1016/j.pragma.2008.03.008
- Vespignani, F., Canal, P., Molinaro, N., Fonda, S., and Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *J. Cogn. Neurosci.* 22, 1682–1700. doi: 10.1162/jocn.2009.21293

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 March 2014; accepted: 05 July 2014; published online: 25 July 2014.

Citation: Kulakova E, Freunberger D and Roehm D (2014) Marking the counterfactual: ERP evidence for pragmatic processing of German subjunctives. *Front. Hum. Neurosci.* 8:548. doi: 10.3389/fnhum.2014.00548

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Kulakova, Freunberger and Roehm. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.