Title
How do iconic co-speech gestures contribute to the truth-conditions of assertions: A surprisal-based ERP investigation targeting N400 and late positivity effects

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Werning, Markus
Reimer, Ludmila
Wieder, Thomas
et al.

Publication Date
2026

How do iconic co-speech gestures contribute to the truth-conditions of assertions: A surprisal-based ERP investigation targeting N400 and late positivity effects

Markus Werning (markus.werning@rub.de)[1], Ludmila Reimer[1], Thomas Wieder[1], Carla Zenk[1] & Maria Spychalska[1]

[1]Department of Philosophy II, Ruhr University Bochum, Universitätsstraße 150, 44870 Bochum, Germany

Abstract

To investigate how co-speech gestures modulate linguistic understanding in comparison to specific verbs with similar contents, we conducted an EEG experiment to measure semantic surprisal by exploring the amplitude changes in the N400 component for both information types. We used videos of a person uttering (i) specified sentences, whose verb semantically denoted a specific action, and (ii) underspecified sentences with an unspecific verb accompanied by an iconic co-speech gesture that indicated the specific action of the previous condition. The subsequent sentence contained an instrument noun as target, which either matched or mismatched the specific action. We measured ERPs on the target noun for both information types and found an N400 effect for mismatching target nouns as well as a late positivity effect for gesture conditions. Crucially, no interaction between information type and congruency was observed in the N400 time-window, indicating similar semantic surprisal induced by mismatching specific verbs and content-similar gestures. An interaction was revealed in the late time-window, suggesting that unlike for gestures, the mismatch negativity effect was prolonged for specific verbs.

Keywords: iconic co-speech gestures; sentence meaning composition; enacted cognition; N400; EEG; semantic surprisal

Introduction

Everyday human communication is multimodal; we employ spoken language, facial expressions, and various gestures to express ourselves. The various sources of information seem to be smoothly understood by conversation partners, yet it is not fully uncovered, how they are integrated on a semantic level. On a cognitive level, gestures interact with speech in many ways: they facilitate language comprehension, production, and acquisition (Holler, 2022; Iverson & Goldin-Meadow, 2005; Özyürek, 2014); they are found to prime or activate internal action representations (Goldin-Meadow & Beilock, 2010; Krauss et al., 2000; Pouw et al., 2014). According to McNeill's (McNeill, 1992) analysis of gesture types and gesture phases, iconic co-speech gestures – gestures accompanying speech and resembling actions or outlining the shape of an object – carry content irreducible to the meanings of the accompanied linguistic expressions. In his analysis, linguistic utterances and iconic co-speech gestures convey information in two parallel streams. Given the theoretical assumption that iconic co-speech gestures have a content similar to the meanings of single words and phrases, the question arises how the content of co-speech gestures interacts with the meanings of words in the successive composition of an assertion's truth-conditions, where we take an assertion to be the combination of the linguistic utterance and the iconic co-speech gesture.

The guiding principle of semantic composition is the principle of compositionality according to which the meaning of a syntactically complex expression is a syntax-dependent function of the meanings of its syntactic parts (Werning, 2005). Now, arguably co-speech gestures – unlike pro-speech gestures – are not an overt syntactic part of the linguistic utterance they accompany. This has led to three rival accounts of how gestural content interacts with the truth-conditions of an assertion: (i) The two-stream view; (ii) the non-at-issue-as-supplement view (Ebert, 2014a, 2014b); and (iii) the non-at-issue-as-cosupposition view (Schlenker, 2018).

(i) The two-stream view assumes that the processing of linguistic meaning is independent of the processing of gestural content. There is no immediate semantic interaction between linguistic meaning and gestural content. In particular, gestural content does not contribute to the truth-conditions of the assertion.[1]

According to the non-at-issue views (ii) and (iii), gestural content does compositionally contribute to truth-conditions of the assertion, however, it only contributes non-at-issue content unless the sentence makes explicit reference to the co-speech gesture, e.g., by a manner demonstrative such as: *NP did it so* where *so* demonstratively refers to the gesture. Non-at-issue content refers to those aspects of meaning that are secondary and not central to the expression. Although they are conveyed along with an expression, they are not the subject of primary negotiation in discourse (Potts, 2007a, 2007b). However, Ebert (2014a, 2014b) and Schlenker (2018) give different analyses of this phenomenon:

(ii) The non-at-issue-as-*supplement* view holds: When iconic co-speech gestures add non-at-issue content, they do so in the manner of supplements similar to appositives and non-restrictive relative clauses: They project into downward-entailing context such as negation and they are non-deniable. Non-deniability means that the assertion cannot be directly denied. Instead, rejecting the assertion requires more complex linguistic means as illustrated in (1) – BIG indicates the content of the iconic co-speech gesture accompanying the expression in square brackets:

[1]Even though McNeill (1992) argues for two independent information streams, it remains unclear whether his analysis would support this view.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

(1) My friend Sophie got a new [dog].
BIG

- # *No, that's not true. Her dog is tiny.*
- $\rightarrow$ *Hey, wait a minute! Actually, her dog is tiny.*

(iii) The non-at-issue-as-cosupposition view agrees with the previous account insofar as iconic co-speech gestures, by default, are regarded as non-at-issue. However, it is now proposed that co-speech gestures behave similar to presuppositions and, since occurring simultaneously with the expression, are called cosuppositions. The idea is that when an expression p co-occurs with a gesture g – with content G – it is asserted that, in the local context, p *entails* G. The gestural content thus projects as presuppositional material. In contrast to the supplement view, the cosupposition view assumes that the gesturally provided content does not merely add information, but semantically restricts the local meaning contribution of the accompanied expression. (1) is hence interpreted as:

- *Sophie got a new dog, and if she got a new dog, it was big.*

Instead, Schlenker (2018) argues, the supplement view works better for post-speech gestures.

In our study, we empirically test the three views against each other by focusing on the semantic surprisal that a mismatching vs a matching discourse continuation invokes. The attribution of the (mis)matching discourse element – a tool used for the action – explicitly references the foregoing discourse element – in our case an action expressed by a verb – where that expression, in one condition, was accompanied by an iconic co-speech gesture (see Figure 1). In a previous study (Reimer et al., 2024), we already found that iconic co-speech gestures do invoke semantic surprisal in the mismatching vs the matching condition. However, in the present study, we want to investigate whether an iconic co-speech gesture accompanying a relatively unspecific (neutral) verb makes about the same semantic contribution to the assertion's truth-conditions as a specific verb (without gesture) whose meaning is approximately identical to the gestural content. We here exploit the experimentally well-established correlation between the semantic surprisal $-log(P(w|c))$ and the size of the N400 component in the event related potential (ERP) measured on the critical word w following a context c in EEG – P is a predictive probability function of the listener (Frank et al., 2013; Kuperberg & Jaeger, 2016; Werning et al., 2019):

$$N400(w|c) \cong -log(P(w|c)).$$

For the three views above, we make the following predictions with regard to semantic surprisal for the the co-speech gesture between the mismatching vs the matching discourse continuation:

(a) If the two-stream view is correct, then there should be only a small or no semantic surprisal effect. For, as hypothesized, the linguistic meaning and the gestural content are processed in independent information streams and gestural content does not contribute to the truth-conditions of the assertion.

(b) If the non-at-issue-as-supplement view is correct, there should be some surprisal effect since the gesture as a non-at-issue supplement does make a semantic contribution to the truth-conditions of the assertion. However, the surprisal effect should be lower than in the case of a content-similar specific verb (without gesture). For, the gesture supplement does not restrict the denoted action, especially, not in a way the content-similar specific verb would.

(c) If the non-at-issue-as-cosupposition view is correct, there should be some surprisal effect and the surprisal effect should be roughly as big as for the content-similar specific verb. For, the gestures's co-supposed entailment restricts the denotation of the action in a way similar to meaning of the content-similar specific verb.

We assume that gestures, even though rather ambiguous on their own, contain information that can be interpreted in a very specific way when combined with speech in a given discourse context. They can add more nuance to the shape of a spoken about object, add details to a described action (such as speed, orientation), and even indicate which tool was used during an action. We are interested in this latter kind, i.e., iconic co-speech gestures that accompany action verbs and provide detailed information on actions and tools used in comparison to the spoken utterances on their own.

N400 & Gestures Predictive processing is widely acknowledged in cognitive and neuroscience as a general mechanism by which the subject at every point in time generates the most probable prediction of the next event on the basis of ongoing perceptual input and learned statistical regularities (Clark, 2013; Hohwy, 2013). To directly investigate whether gestures modulate the expectations of listeners, we will use event-related brain potentials (ERPs), focusing especially on the N400 component. ERPs are scalp-recorded voltage changes time-locked to trigger events, such as spoken or written word. The N400 is a negative shift in the ERP waveform starting at ca. 200ms and peaking around 400ms post-stimulus onset over the centro-parietal scalp sites (Kutas & Federmeier, 2011; Swaab et al., 2011). It is modulated by the meaning-related predictability of the stimulus, e.g., it tends to be larger for words that are semantically less appropriate or less expected in the context (Kutas & Federmeier, 2000; Kutas & Hillyard, 1980; Kutas & Van Petten, 1994). The size of the N400 is inversely correlated with the cloze probability of the triggering word, which is measured as the percentage of individuals who would continue a given sentence fragment with that word (Federmeier et al., 2007). Prior EEG research has shown that gestures can modify the N400 component or even elicit a similar effect themselves: In such studies, gestures elicited an N400-like effect when they did not match a preceding video or picture, referred to as N450 (Wu & Coulson, 2005). Moreover, words as well as pictures elicit an N400 effect when they do not match a preceding (silent) gesture, suggesting that

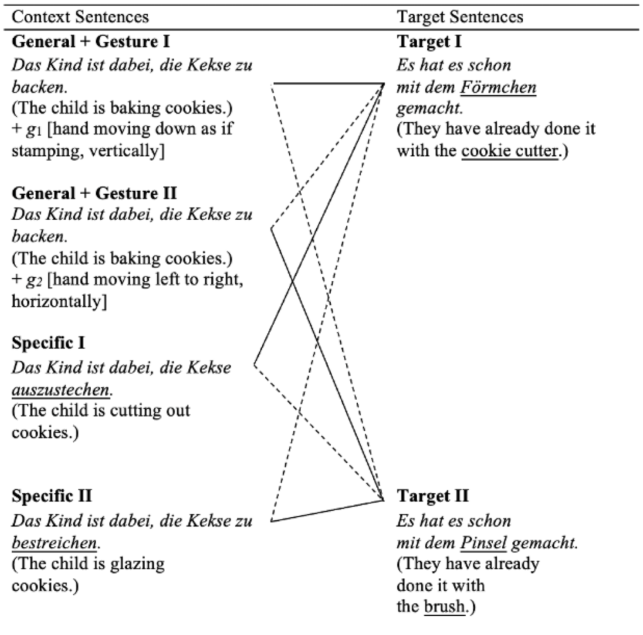

Figure 1: Experimental design: Solid lines indicate congruent continuations (matches), dashed lines incongruent continuations (mismatches). Cue and target words are underscored.

gestures can establish expectations about upcoming linguistic input (Wu & Coulson, 2007, 2015). As described above, the N400-effect has been linked to the computation of semantic content; thus, it can be assumed that gestures at least interact with meaning and possibly carry meaning themselves. In another ERP study, it was shown that gestures help to disambiguate homonyms and that mismatching gesture elicit N400 effects (Holle & Gunter, 2007). However, the same paper reports that the presence of "meaningless gestures", i.e., grooming hand movements, reduce the N400 effects, and also the overall helpfulness of the iconic gestures to disambiguate the homonym. (Obermeier et al., 2011) conducted follow up studies, using gesture fragments and manipulated the task and the synchrony of gesture and speech. Their experiments showed that if speech and gesture were off-set and paired with a task that required participants explicitly rating the compatibility of gesture and speech, the effect was similar to a set-up with no overt task but synchronous speech and gesture. However, in another variation without an overt task nor an asynchrony, an N400 could not be detected. This again is in favor of multiple factors that influence how gestures are processed with speech (or how speech is processed in the presence of gestures).

Method

The goal of this study is to measure semantic surprisal using the N400 component of the ERP and thereby investigate whether and to which degree co-speech gestures modulate truth-conditions of assertions. We compare iconic co-speech gestures accompanying an unspecific (neutral) verb to specific action verbs whose meaning is approximately identical to the gestural content in a given discourse context. In our 2x2 design *Information Type* (gesture, specific verb) x *Congruency* (match, mismatch), we use naturalistic videos of a person either uttering 1) underspecified action sentences, accompanied by iconic co-speech gestures that represent a more specific action, or 2) specified action sentences, featuring a content-similar specific action verb, without any gestures. The videos were followed by target sentences containing an instrument noun as the target word. The noun either matched or mismatched the previously seen gesture and, respectively, the specific action verb. We used carefully constructed materials in which the linguistic material including the neutral verb did not vary in regard to modulating predictions about upcoming words, so any change in the processing would be due to a change in the gesture for the underspecified context sentences. This approach is novel in two regards: First, experiments measuring ERPs elicited by linguistic targets following a gesture used paradigms of isolated, silent gesture videos, and used single probe words. The linguistic targets elicited N400-like effects, however, this depended on the timing between gesture and probe word (Habets et al., 2011; Wu & Coulson, 2007). In contrast, our sentences are part of the same discourse and do not rely solely on the simultaneity of gesture and speech. Second, other experiments using full sentences focused on the (mis)match between a co-speech gesture and a simultaneously uttered phrase or word (Hintz et al., 2023; Özyürek et al., 2007), not on the mismatch elicited by a gesture in regard to the following linguistic input, as we do. Some of these studies also varied the linguistic context sentences, adding another factor that might influence the processing of both utterances and co-speech gestures (Hintz et al., 2023), while we kept linguistic input constant and varied only the gestures.

Design

We used a 2 x 2 design with two factors: 1) *Congruency* (mismatch, match), and 2) *Information Type* (gesture, specific verb). Information Type depended on the context videos and Congruency on the displayed target sentence.

In the gesture conditions, the context sentences contained an unspecific (neutral) verb v_{gen} accompanied by a co-speech gesture g shown in the video as cue and a noun y, for which the verb denotes an appropriate action:

$$\text{The agent } x \text{ is } v_{gen}\text{-ing object } y.$$

In the specific-verb condition, the context sentences contained a specific verb v_{spec} and the same object noun y:

$$\text{The agent } x \text{ is } v_{spec}\text{-ing object } y.$$

Target sentences contained the target noun n, which either matches or mismatches the specific verb v_{spec} and, respectively, the content-similar gesture g:

$$\text{The agent } x \text{ has already done it with } n.$$

Note that, in the German stimuli, all target sentences only divert in the end. An overview of the stimulus design is shown in Figure 1.

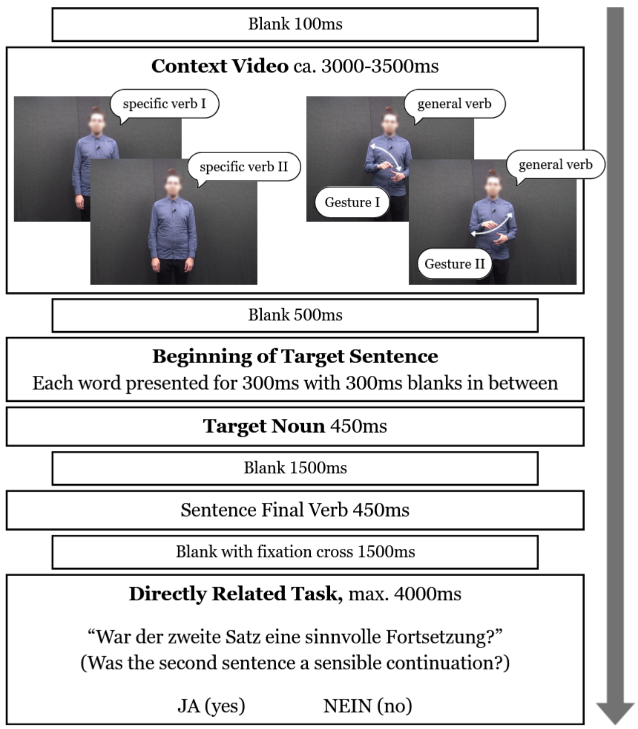

Figure 2: Schematic representation of one trial.

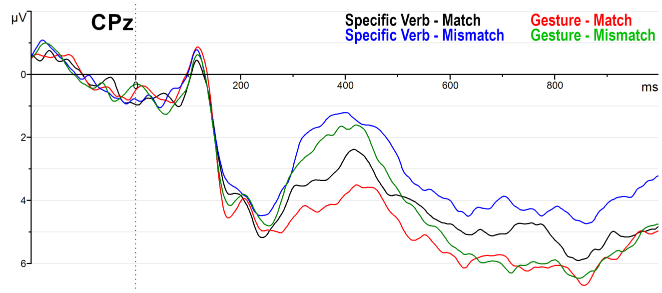

Figure 3: Grand averages for target nouns in CPz.

Material Preparation

All pairs of cue verbs and target nouns were controlled for frequency and for the grammatical gender of nouns n. To ensure that both target sentences (in the matching and mismatching conditions) were equally likely to follow context sentence containing the unspecific verb, we controlled for the GloVe values (Pennington et al., 2014) between: (1) context sentences and the unspecific cue verbs v_{gen}, (2) context sentences and the nouns n, and (3) the cue verb v_{gen} and the target noun n. Regarding *congruency*, we made sure that the GloVe values between the specific verb v_{spec} and the target noun were high ($> .15$) in the matching and low ($< .15$) in the mismatching condition.

Each set of video material for the current experiment consisted of a) two videos with no gestures displaying spoken sentences with the specific verbs v_{spec}, and b) two videos displaying spoken sentences with the unspecific verb v_{gen} accompanied by an iconic co-speech gesture g, which corresponds to the actions expressed by the respective specific verbs. We built on a pool of videos from an earlier study (Reimer et al., 2024).

All videos were recorded in full HD resolution (1920x1080 pixel) and a frame rate of 50fps. The speaker was recorded from the knees up, allowing for framing with the hands to be visible at all times. The videos were cut with ca. 600ms (30 frames) before gesture or speech onset and 600ms after offset were left. Speech on- and offsets were determined by inspecting the audio track, and gesture on- and offsets were defined by the hands leaving and returning to their resting position, see Figure 2. All of our material in the pool was constructed with uniform constraints and was later assessed the suitability of the material in online surveys.

To ensure similarity between the content of a gesture g and the meaning of the corresponding specific verb v_{spec}, we had a separate set of participants rate how well specific verbs described the gestures shown in the videos. Overall, only gestures whose content similarity received high ratings were selected for the experiment.

We supplemented the earlier produced material by recording new audio material with the same speaker, and ensured that the new audio material matched the previous one in tone, cadence, and speed. The new audio was overlaid on gestureless videos from our previous video pool. We blurred the face of our speaker in the videos and the length of the sentences in syllables remained the same. For further details on video recording, post-processing, and stimulus constraints, see Reimer and Werning (2023) and Reimer et al. (2024). An example set of our material can be seen in Figure 2.

Participants and Procedure

We recruited a total of 40 participants (25 female, age range 20-33, mean age 25.18, SD 3.80). All were right-handed, with normal or corrected-to-normal vision, had no neurological or psychological impairments, and were monolingual German native speakers. Four participants were excluded due to technical issues and three were excluded from the analysis due to the insufficient number of usable segments. The final analyzed sample contained 33 subjects (21 female, aged 20-32, mean age 24.70, SD 3.51).

The experiment was conducted in the EEG laboratory of the Chair of Language and Cognition, Ruhr University Bochum, which has been approved by an ethics statement of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS).

Participants were seated inside an electrically isolated and acoustically attenuated cabin, in front of a shielded glass with a computer screen behind it. USB-powered speakers were placed inside the cabin as well as a Cedrus response pad with two designated buttons. Every subject signed a written consent of participation and was informed that their data will be stored and handled fully anonymously, and that they have the right to withdraw from the experiment at any time. Participants were screened for their Autism-Spectrum Quotient

(Baron-Cohen et al., 2001) and working memory.

The experiment started with an instruction screen and six exercise trials with feedback. The trials of the main experiment did not provide feedback. The experiment was split into 10 blocks with 32 trials each, 320 trials total. One block lasted ca. 8 min, and the session took about 90 min, breaks included. Each trial started with a video of the speaker uttering the context sentence followed by the target sentence word by word. After a blank screen, participants were prompted to evaluate whether the displayed sentence was a sensible continuation of the video (see Figure 2). They answered by pressing a button corresponding to "Yes" or "No", which were displayed on pseudo-randomly alternating sides, "Yes" being left 50% of the time.

The EEG was recorded using a 64-channel ActiCAP snap system (Brain Products), with 60 active Ag/AgCl electrodes placed according to the 10-20 system. FCz location was used as the reference and AFz as the ground electrode. Four additional electrodes recorded eye movements: two were placed above and below the right eye to monitor vertical movements, and two at the temples to monitor horizontal movements. Impedance was kept below 5 kΩ. Recording was done using Brain Vision Recorder, with a sampling rate of 1000 Hz, a resolution of 0.1 µV, a low cut-off time constant of $10s$ (≈ 0.016 Hz), and a high cut-off filter of 250 Hz. The EEG data was processed with Brain Vision Analyzer 2.2.

Breaks and large artifacts were excluded manually from the data. The data was down-sampled to 500 Hz. An offline band-pass zero-phase (forward–backward) 4th-order Butterworth filter of 0.1- 30 Hz was applied. Automatic raw data inspection rejected all periods where signal changed more than $150\,\mu$V/150 ms or less than $0.5\,\mu$V per 100 ms intervals. The maximal voltage step was $50\,\mu$V/ms. Electrodes recording eye activity and anterior channels were excluded from this procedure to avoid classifying blinks as artifacts. Then, both vertical and horizontal eye-movements were corrected by means of independent component analysis. Data was re-referenced to the average of mastoid electrodes (TP9 and TP10). Segments from 200 ms pre-target onset until 1000 ms post-onset were extracted for every trial and condition. Baseline correction used the 200 ms interval preceding the stimulus onset. Segments with any remaining physical artifacts, including those with the amplitude lower than $-90\,\mu$V or higher than $90\,\mu$V, were excluded and condition averages were calculated for each subject. Data sets with less than 60% segments in any condition were excluded, remaining subjects had 28 to 40 segments (mean 37.47, SD 2.41).

Statistical analyses

Statistical analyses were performed in R (Version 2026.01.0+392). We used the *tidyverse* suite of packages for data wrangling and visualization. ERP amplitudes were averaged across the region of interest (ROI) including electrodes Cz, CPz, and Pz for two time-windows: the N400 (300-500ms) and late time-window (600-800 ms). Two 2 x 2 repeated-measures ANOVA were conducted using the *afex* package, where Information Type and Congruency were used as within-subject factors. All models used Type 3 Sums of Squares. Post-hoc pairwise comparisons were performed using the *emmeans* package. To control for multiple comparisons, p-values were adjusted using the Bonferroni correction. Effect sizes for pairwise differences were calculated as Cohen's d_z for within-subject designs ($d_z = \frac{t}{\sqrt{n}}$). Data visualizations were generated using *ggplot2* (via *tidyverse*) and the *ggsignif* package.

Results

Continuity Judgements Participants performed the task well, they accepted the matching targets with 84.38% mean accuracy (SD 14,49), i.e. accepted the continuation, in the gesture contexts and with 88.81% (SD 12.98) in the specific verb contexts. They performed worse for rejecting the mismatching targets, with 71.88% (SD 23.67) accuracy for the gesture and 76.98% (SD 10.84) for the verb condition. As the main focus of this paper is the EEG analyses, we do not report further analyses on the behavioral data.

N400 In the N400 time-window, there was a significant main effect of Information Type, F(1, 32) = 12.39, p = .001, η^2_G = .009, with less negative amplitudes for gesture (M = 3.14, SE = 0.468) than for specific verb (M = 2.47, SE = 0.406). A significant main effect of Congruency was also found, F(1, 32) = 73.37, p < .001, η^2_G = .049, with less negative amplitudes for match (M = 3.59, SE = 0.436) compared to mismatch (M = 2.02, SE = 0.423). The interaction between Information Type and Congruency was not significant: F(1, 32) = 0.02, p = .888, η^2_G < .001, see Figure 4.

Late time-window There was a significant main effect of Information Type, F(1, 32) = 29.84, p < .001, η^2_G = .043: gestures elicited more positive amplitudes (M = 5.765) than specific verbs (M = 4.505). The main effect of Congruency was not significant, F(1, 32) = 1.27, p = .269, η^2_G = .002. A significant interaction was found between Information Type and Congruency, F(1, 32) = 4.33, p = .045, η^2_G = .006: The difference between gesture and specific verb was larger for mismatch ($\Delta = 1.70\,\mu V$) than for match ($\Delta = 0.82\,\mu V$). Bonferroni corrected pairwise comparisons showed that for specific verbs, matches (M = 4.84, SE = 0.495) elicited significantly larger late positivity amplitudes than mismatches (M = 4.17, SE = 0.544), t(32) = 2.37, p = .0238, d_z = 0.41, but there was no significant difference between matches (M = 5.66, SE = 0.579) and mismatches (M = 5.87, SE = 0.466), for gestures t(32) = -0.69, p = .4953, d_z = -0.12, see Figure 4.

Discussion

Congruency We observed a robust mismatch-match N400 effect for both information types (gesture, specific action verb), where mismatching contexts elicited larger negativities in the 300-500 ms time window. Taking the N400 as a measure of semantic surprisal, this result does not support the

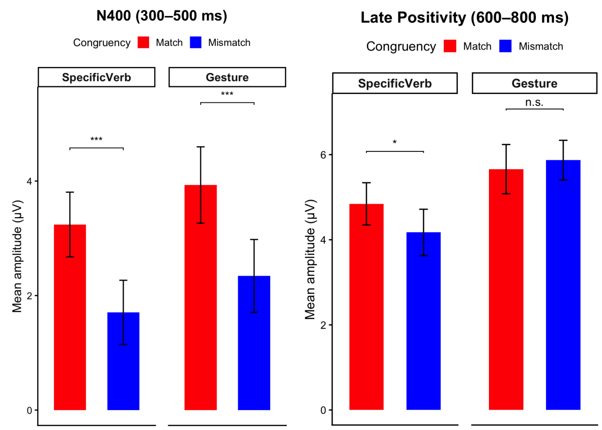

Figure 4: Left: Mean N400 amplitudes (300–500 ms) at the centro-parietal ROI (Cz, CPz, Pz) for matching and mismatching stimuli, separately for specific verbs and gestures.
Right: Mean late positivity amplitudes (600–800 ms) at the centro-parietal ROI (Cz, CPz, Pz) for matching and mismatching stimuli, separately for specific verbs and gestures.

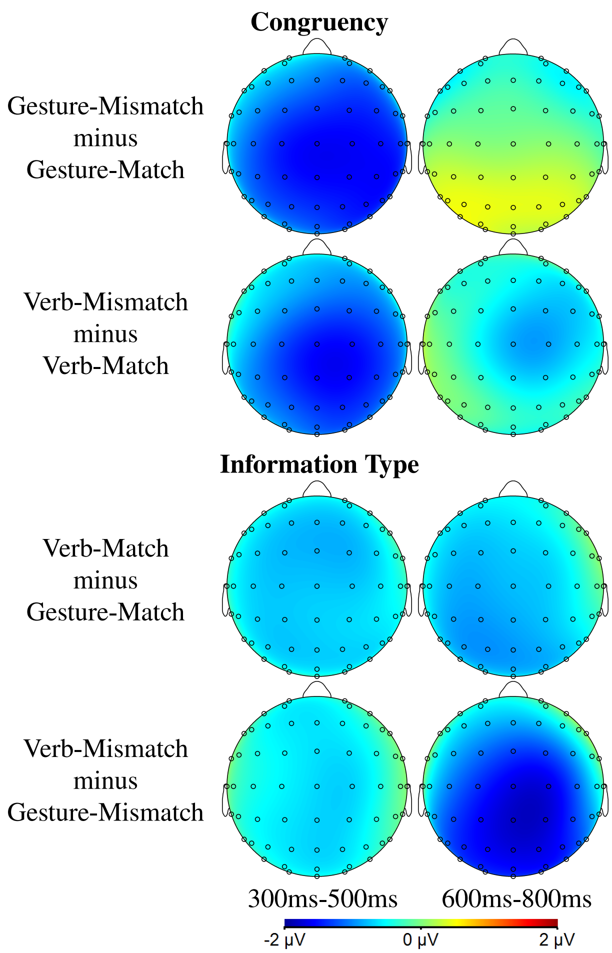

Figure 5: Topographical plots showing the differences between the conditions.

two-stream view, which predicts that there is only a small or no semantic surprisal effect for the co-speech gesture between the mismatching and the matching discourse continuation. For, as hypothesized by this view, the linguistic meaning and the gestural content are processed in independent information streams and gestural content does not contribute to the truth-conditions of the assertion.

Crucially, congruency effects did not differ significantly between information types, with both conditions, gesture and specific verb, showing large effect sizes. This result is consistent with the not-at-issue-as-cosupposition view, which predicts that the surprisal effect between the Mismatch and Match condition for the co-speech gesture should be roughly as big as for the content-similar specific verb. The reason is that the gesture's co-supposed entailment restricts the denotation of the action in approximately just the way the meaning of the content-similar specific verb does. The N400 results, in contrast, do not support the non-at-issue-as-supplement view. Here, the surprisal effect for the gesture should be lower than in the case of the content-similar specific verb. For, the gesture supplement – as this view hypothesizes – does not restrict the denoted action, especially, not in a way the content-similar specific verb would.

Information Type Late effects, however, revealed a qualitative difference between information types in that gestures led to a larger late positivity relative to specific verbs. In the current literature, late positivities are associated with increased demands for semantic integration, reanalysis, and the updating of situation models following initial lexical access (Brouwer et al., 2012). In addition, the interaction between Information type and Congruency showed that the congruency effect differed reliably between gestures and specific verbs: a significant late mismatch-match effect was observed for specific action verbs but not for gestures. Given that we observe a smaller positivity for linguistic but not for gestural mismatches relative to matches in the late time window, this effect could be interpreted as a prolonged negativity. Such prolonged negativities have been observed in cases requiring more effortful inferential and contextual processing (Baggio et al., 2008; van Berkum et al., 2003, 2007).

Conclusion Overall, our findings suggest functional differences between early and late stages of semantic processing in the light of content mismatches, comparing the compositional contribution of specific verbs and content-similar iconic co-speech gestures. Early semantic processes seems to operate similarly across both information types, while later semantic processing is different between information types. Regarding the contribution of iconic co-speech gestures to the truth-conditions of assertions, the view that gestures are non-at-issue cosuppositions is supported by our study, whereas alternative views such as the two-stream view and the non-at-issue-as-supplement view are not.

References

Baggio, G., van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An erp study. *Journal of Memory and Language*, *59*(1), 36–53. https://doi.org/10.1016/j.jml.2008.02.005

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (aq). *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the p600 in language comprehension. *Brain Research*, *1446*, 127–143. https://doi.org/10.1016/j.brainres.2012.01.055

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–253.

Ebert, C. (2014a). The non-at-issue contributions of gestures and speculations about their origin. *Workshop Demonstration and Demonstratives*.

Ebert, C. (2014b). The semantic impact of co-speech gestures. In *Embodied meaning goes public*.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2013). Word surprisal predicts n400 amplitude during reading [citation-key: frank-etal-2013-word]. *Proceedings of the 51st annual meeting of the association for computational linguistics (volume 2: Short papers)*, 878–883. https://aclanthology.org/P13-2152/

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, *5*(6), 664–674.

Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration. *Journal of Cognitive Neuroscience*, *23*(8), 1845–1854. https://doi.org/10.1162/jocn.2010.21462

Hintz, F., Khoe, Y.-H., Strauß, A., Psomakas, A. J. A., & Holler, J. (2023). Electrophysiological evidence for the enhancement of gesture-speech integration. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(2), 340–353. https://doi.org/10.3758/s13415-023-01074-8

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.2007.19.7.1175

Holler, J. (2022). Visual bodily signals as core devices for coordinating minds. *Philosophical Transactions of the Royal Society B*, *377*(1859), 20210094. https://doi.org/10.1098/rstb.2021.0094

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371.

Krauss, R. M., Chen, Y., & Gotfexnum, R. F. (2000). 13 Lexical gestures and lexical access: A process model. *Language and gesture*, *2*, 261.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language Cognition & Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

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*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences. *Science*, *207*(4427), 203–205.

Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In *Handbook of psycholinguistics* (pp. 83–143).

McNeill. (1992). Guide to Gesture Classification, Transcription and Distribution. *Hand and Mind: What Gestures Reveal about Thought*. https://doi.org/10.2307/1576015

Obermeier, C., Holle, H., & Gunter, T. C. (2011). What iconic gesture fragments reveal about gesture-speech integration: When synchrony is lost, memory can help. *Journal of Cognitive Neuroscience*, *23*(7), 1648–1663. https://doi.org/10.1162/jocn.2010.21498

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130296. https://doi.org/10.1098/rstb.2013.0296

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605–616. https://doi.org/10.1162/jocn.2007.19.4.605

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Potts, C. (2007a). Into the conventional-implicature dimension. *Philosophy Compass*, *2*(4), 665–679. https://doi.org/https://doi.org/10.1111/j.1747-9991.2007.00089.x

Potts, C. (2007b). *The Logic of Conventional Implicatures*. https://doi.org/10.1093/acprof:oso/9780199273829.001.0001

Pouw, W. T., de Nooijer, J. A., van Gog, T., Zwaan, R. A., & Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, *5*(APR), 1–14. https://doi.org/10.3389/fpsyg.2014.00359

Reimer, L., Spychalska, M., & Werning, M. (2024). Integrating Co Speech Gestures into Sentence Meaning Comprehension. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*, 5025–5032.

Reimer, L., & Werning, M. (2023). Modelling the Integration of Co-Speech Gestures into Sentence Meaning Composition. In M. Goldwater, F. Anggoro, B. K. Hayes, & D. Ong (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 625–631, Vol. 45).

Schlenker, P. (2018). Gesture projection and cosuppositions. *Linguistics and Philosophy*, *41*(3), 295–365.

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2011). Language-Related ERP Components. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0197

van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology*, *40*(2), 235–248. https://doi.org/10.1111/1469-8986.00025

van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, *1146*, 158–171. https://doi.org/10.1016/j.brainres.2006.06.091

Werning, M. (2005). Right and wrong reasons for compositionality. In M. Werning, E. Machery, & G. Schurz (Eds.), *The Compositionality of Meaning and Content* (pp. 285–309, Vol. 1). Ontos Verlag.

Werning, M., Unterhuber, M., & Wiedemann, G. (2019). Bayesian Pragmatics Provides the Best Quantitative Model of Context Effects on Word Meaning in EEG and Cloze Data. In A. Goel, C. Seifert, & C. Freska (Eds.), *Proceedings of the 41th Annual Conference of the Cognitive Science Society* (pp. 3085–3091). Cognitive Science Society.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.2005.00356.x

Wu, Y. C., & Coulson, S. (2007). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03194028

Wu, Y. C., & Coulson, S. (2015). Iconic Gestures Facilitate Discourse Comprehension in Individuals With Superior Immediate Memory for Body Configurations. *Psychological Science*, *26*(11), 1717–1727. https://doi.org/10.1177/0956797615597671

Acknowledgments

On the side of the PI MW, this project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project numbers GRK-2185/2 (DFG Research Training Group Situated Cognition); 367110651 (within the Priority Program SPP 1727); 579381973 (within the Priority Program SPP 2556).