

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

Anonymous Submission

## 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 (1992) analysis of gesture types and gesture phases, iconic co-speech gestures – gestures accompanying speech and resembling concrete 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 semantic content similar to the meanings of single words and phrases, the question arises how the semantic 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.<sup>1</sup>

According to the non-at-issue views (ii) and (iii), gestural content does compositionally contribute to truth-conditions

<sup>1</sup>Even though McNeill (1992) argues for two independent information streams, it remains unclear whether his analysis would support the this view.

of the assertion, however, it only contributes non-at-issue content (Potts, 2004) 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. However, Ebert and Schlenker 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 illustrate in (1) – BIG indicates the content of the iconic co-speech gesture accompanying the expression in square brackets:

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

BIG

- # *No, that's not true. Her dog is tiny.*
- → *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 Fig. 1). In a previous study

(Reimer, Spychalska, & Werning, 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 (Werning, Unterhuber, & Wiedemann, 2019; Kuperberg & Jaeger, 2016; Frank, Otten, Galli, & Vigliocco, 2013):

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

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 the 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.

**Design** We used a 2 x 2 design with the variables: 1) *Congruency* (mismatch, match), and 2) *Information Type* (gesture, specific verb) – see Fig. 1. Our context sentences had the following structure: In the gesture condition, the context sentence 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:

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

In the specific-verb condition, the context sentence contained an specific verb  $v_{spec}$  as cue and the same noun  $y$ :

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

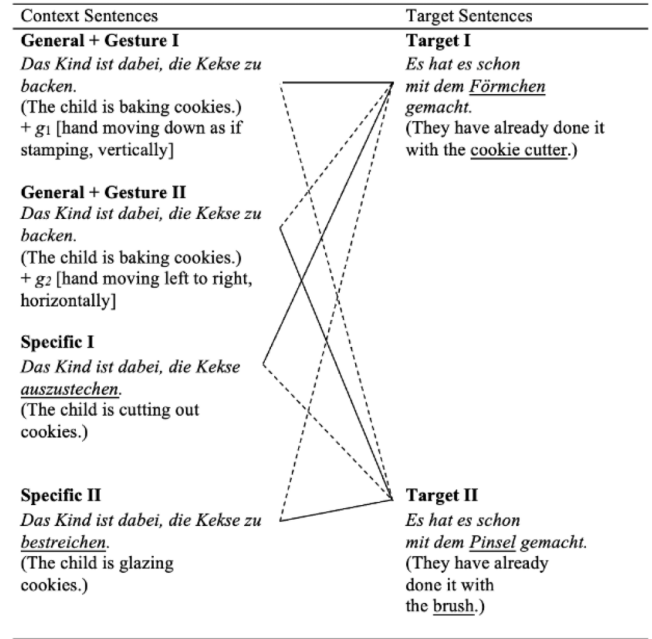


Figure 1: Experimental design: Solid lines indicate congruent continuations (matches), dashed lines incongruent continuations (mismatches). Cue and target words are underscored.

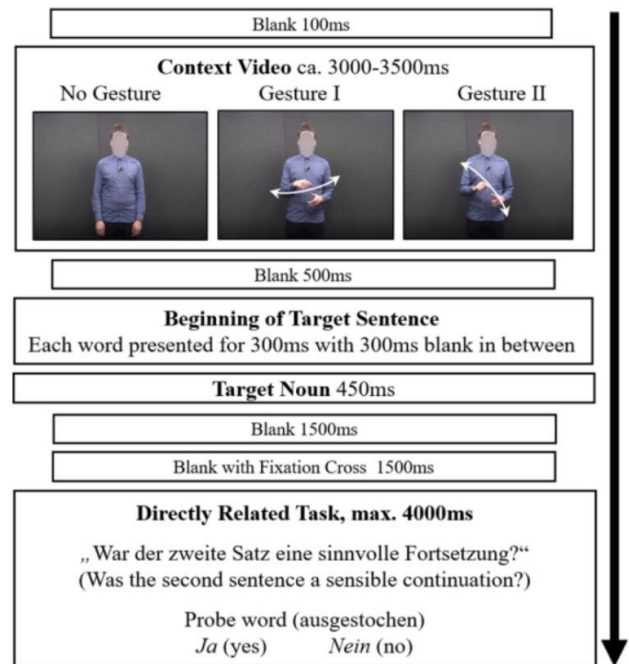


Figure 2: Schematic representation of one trial.

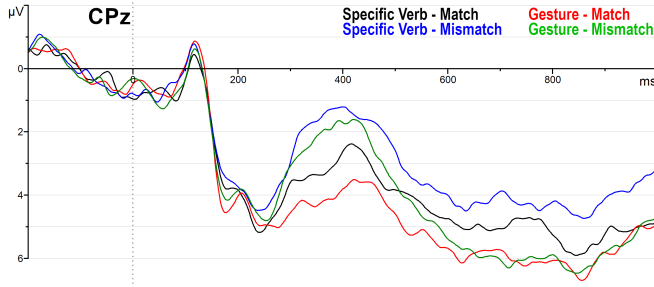


Figure 3: Grand averages for target nouns in CPz.

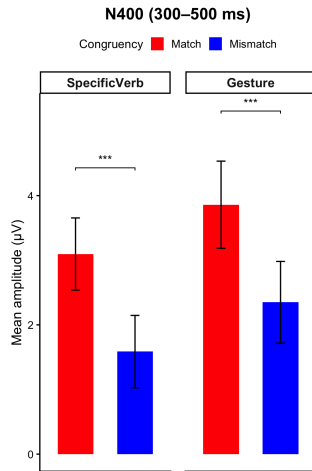


Figure 4: Mean N400 amplitudes (300–500 ms) at the centro-parietal ROI (Cz, CPz, Pz) for matching and mismatching stimuli, separately for specific verbs and gestures.

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

The agent  $x$  has already done it with  $n$ .

For the schematic presentation of one trial see Fig. 2.

**Participants and Procedure** We recruited a total of 37 German speaking participants (24 female, age range 18-40, mean age 24.97, SD 3.84), right-handed, without neurological, psychological or visual impairments. Four were excluded due to technical difficulties. An ethics approval of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS) was obtained.

## Discussion

**N400** The N400 results – see Fig. 3 & 4 – showed a robust mismatch-match 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 for semantic surprisal, this result disconfirms the two-stream view which predicts that there is only a

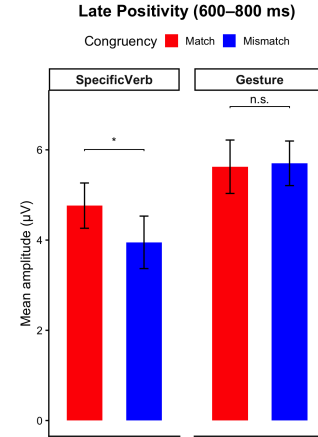


Figure 5: Mean late positivity amplitudes (600–800 ms) at the centro-parietal ROI (Cz, CPz, Pz) for matching and mismatching stimuli, separately for specific verbs and gestures.

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 supports the not-at-issue-as-cosupposition view. For, this view predicts that there is some surprisal effect for the co-speech gesture between the mismatching and the matching discourse continuation, and this surprisal effect 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.

**Late positivity** Late positivity effects – see Fig. 3 & 5 – however, revealed a qualitative difference between information types. A significant mismatch-match effect was only observed for specific action verbs. Gestures, on the other hand, did not elicit a corresponding late positivity. Additionally, the interaction between Information Type and Congruency indicates that the congruency effect differed reliably between gestural and linguistic stimuli. In the current literature, late positivities are associated with post-semantic processes. Our results suggest that linguistic mismatches may trigger additional evaluative repair-processes, unlike gestural mismatches. Possible explanations include stronger demands on explicit semantic revision when expectations are violated.

The absence for gestural contexts may also be explained by a greater tolerance for ambiguity in gestural meaning. Since gestures convey information in a context-dependent manner, they may reduce the need for explicit reanalysis when mismatches occur.

Overall, our findings suggest functional differences between early and late stages of semantic processing. Early semantic integration seems to operate similarly across both information types, while later semantic processing are different between information types. The results highlight the importance of distinguishing between different stages of semantic processing when investigating multimodal communication. Gestures and language seem to share early integration mechanisms but differ in how those semantic conflicts are resolved at later semantic processing stages.

## References

- Ebert, C. (2014a). The non-at-issue contributions of gestures and speculations about their origin. In *Workshop demonstration and demonstratives*.
- Ebert, C. (2014b). The semantic impact of co-speech gestures. In *Embodied meaning goes public*.
- Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2013, August). Word surprisal predicts n400 amplitude during reading. In H. Schuetze, P. Fung, & M. Poessio (Eds.), *Proceedings of the 51st annual meeting of the association for computational linguistics (volume 2: Short papers)* (p. 878–883). Sofia, Bulgaria: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/P13-2152/> (citation-key: frank-et-al-2013-word)
- Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, 5(6), 664–674.
- Holler, J. (2022). Visual bodily signals as core devices for coordinating minds. *Philosophical Transactions of the Royal Society B*, 377(1859), 20210094. doi: 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., & Gottfexnum, 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. doi: 10.1080/23273798.2015.1102299
- McNeill. (1992). Guide to Gesture Classification, Transcription and Distribution. *Hand and Mind: What Gestures Reveal about Thought*. doi: 10.2307/1576015
- 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. doi: 10.1098/rstb.2013.0296
- Potts, C. (2004). *The logic of conventional implicatures*. Oxford University Press. Retrieved from <https://academic.oup.com/book/368> doi: 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. doi: 10.3389/fpsyg.2014.00359
- Reimer, L., Spsychalska, 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.
- Schlenker, P. (2018). Gesture projection and cosuppositions. *Linguistics and Philosophy*, 41(3), 295–365.
- Werning, M. (2005). Right and wrong reasons for compositionality. In M. Werning, E. Machery, & G. Schurz (Eds.), *The Compositionality of Meaning and Content* (Vol. I, pp. 285–309). Frankfurt: 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 *Proceedings of the annual meeting of the cognitive science society* (Vol. 41, p. 3085-3091).