Multimodal language processing: How preceding discourse constrains gesture interpretation and affects gesture integration when gestures do not synchronise with semantic affiliates☆

Isabella Fritz [a,b,*], Sotaro Kita [c], Jeannette Littlemore [a], Andrea Krott [d]

[a] *Department of English Language and Linguistics, University of Birmingham, Birmingham B15 2TT, UK*
[b] *Norwegian University of Science and Technology (NTNU), Language Acquisition and Languagse Processing Lab, 7491 Trondheim, Norway*
[c] *Department of Psychology, University of Warwick, CV4 7AL Coventry, UK*
[d] *School of Psychology, University of Birmingham, B15 2TT Birmingham, UK*

ABSTRACT

Previous studies have suggested that a co-speech gesture needs to be synchronous with semantically related speech (semantic affiliates) for its successful semantic integration into a discourse model because co-speech gestures are often highly ambiguous on their own. But not all gestures synchronise with semantic affiliates, some precede them. The current study tested whether the interpretation of a gesture that does not synchronise with its semantic affiliate can be constrained by preceding verbal discourse and integrated into a recipient's discourse model. A behavioural experiment (Experiment 1) showed that related discourse information can indeed constrain recipients' interpretations of such gestures. Results from an ERP experiment (Experiment 2) confirmed that synchronisation between gesture and semantic affiliate is not essential in order for the gesture to become part of a discourse model, but only if the preceding context constrains the gesture's meaning. In this case, we found evidence for post-semantic integration (P600, time-locked to the gesture's semantic affiliate).

Introduction

In everyday conversation, information is not only conveyed via speech but also through gestures that accompany speech (Goldin-Meadow, 2003; McNeill, 1992, 2005). In comprehension, meanings conveyed via co-speech gesture and meanings conveyed via speech are automatically combined (Kelly, Özyürek, & Maris, 2010). This gesture-speech integration process is modulated by various factors, including gesture-speech synchrony, i.e. temporal overlap (see Özyürek, 2014 for a review). Synchrony between a gesture and semantically related portion of speech appears to be crucial in order to form a unified representation of a message. This is because gestures are highly ambiguous when they occur without semantically related words in speech (see Kelly, Manning, & Rodak, 2008 for a review). But gestures do not always synchronise with semantically related words they often precede them (Fritz, 2018; Morrel-Samuels & Krauss, 1992; Schegloff, 1984). It is not clear if and how such gestures can be integrated with speech during comprehension.

Speech comprehension depends highly on discourse context (e.g., van Berkum, 2008). In order to construct the meaning of an utterance, we relate incoming information to already established referents (e.g., van Berkum, Koornneef, Otten, & Nieuwland, 2007), use discourse information to predict upcoming words (e.g., Otten & van Berkum, 2008), and integrate incoming information into an existing discourse model (e. g., Brouwer, Fitz, & Hoeks, 2012). The present study explores whether discourse information is used not only to integrate speech, but also to integrate gestural meaning and more specifically the meanings of those gestures that synchronise with semantically unrelated words.

Previous studies on the integration of gestures and speech have often used electroencephalography and event-related potentials (ERPs) (e.g., Habets, Kita, Shao, Özyürek, & Hagoort, 2011; Wu & Coulson, 2010; Holle & Gunter, 2007; Kelly, Kravitz, & Hopkins, 2004; Özyürek, Willems, Kita, & Hagoort, 2007). For instance, Özyürek et al. (2007) found that gestures and/or co-occurring spoken words (*walking* gesture and/or the word "walk") that mismatch the preceding sentence context ("He slips on the roof…") both led to N400 effects, and these N400 effects had

the same onset latency. Since the N400 component is sensitive to semantic processing cost (see Kutas & Federmeier, 2011 for a review), these results suggest that gestures and their co-occurring spoken words are simultaneously integrated into the preceding sentence context and thus become part of a unified representation of the message, that is a recipient's discourse model.

Gestures are not only integrated into a speech context, but also into a non-speech context. Wu & Coulson (2005) found that iconic gestures (presented without speech) that mismatched preceding cartoons elicit an N400 effect compared to matching gestures. Thus, gestures are integrated not only into linguistic but also non-linguistic discourse contexts.

Wu and Coulson (2010) showed participants discourse primes consisting of either speech only (e.g., "Where there is a green parrot – fairly large") or speech-gesture combination (e.g., adding a gesture that indicated the size of the parrot). These primes were followed by either a related picture probe (e.g., parrot) or an unrelated picture probe (e.g., hourglass). They observed two ERP effects: First, a reduced N400 time-locked to the gesture compared to the no-gesture condition indicating a facilitating nature of iconic gestures on lexical retrieval of its semantic affiliate. Second, the amplitude difference in the N400 time-window when comparing the matching vs. mismatching probes was larger in the gesture condition compared to the non-gesture condition suggesting that gestures were integrated into the discourse context.

Furthermore, Holle and Gunter (2007) investigated gesture-speech integration by testing whether an iconic gesture (e.g., a *dancing* gesture) can disambiguate a co-occurring homonym (e.g., "ball" = dance or game in *Sie beherrschte den Ball, was sich im Spiel/Tanz…* "She controlled the ball [= ambiguous noun], which during the game/dance…"), and then this disambiguated concept becomes discourse context. ERPs were time-locked to the target word downstream in the sentence that either matched (e.g., "dance") or mismatched (e.g., "game") the gesture and disambiguated homonym. They found a larger N400 on the target word when the gesture mismatched with the target word compared to when it matched. Thus, the gesture together with the homonym formed an idea unit in the discourse model that modulated sentence processing downstream.

Synchrony between co-speech gesture and semantically related words

McNeill (1992) proposed the "Semantic Synchrony Rule", which states that gestures synchronise, that is temporally fully or partly overlap, with semantically related words in speech. This rule was mainly based on qualitative analysis of English narratives. However, it is not always followed in actual production. As other qualitative analyses have indicated, iconic gestures are sometimes produced and terminated before the semantically related words. For example, Kita (2000) discussed a case in which an adult English speaker who, when describing a complex motion sequence, initially struggled with description and said "because it **aaaa**. Well actually what happens is, he I, you, **assume**". The speaker produced gestures depicting the motion sequence with the filler "aaaa" and the word "assume", before she started to describe the sequence ("that he swallows the bowling ball and comes rolling out"), in which "swallow" and "come rolling out" are the semantically related words. McNeill himself has discussed similar cases. He reported a case in which an adult Turkish speaker produced a gesture depicting manner of motion one phrase before the co-expressive manner word was produced in speech (McNeill, 2005, pp. 138-141; taken from Özyürek, 2001). As the speaker says "the ball in one way [i.e., somehow] while hopping", he produced an iconic gesture for hopping with the phrase "the ball in one way", foreshadowing the semantically related manner verb. He also presented a case where a speaker produced a gesture indicating the location of a character while producing the phrase "flashing back" in the sentence "they keep on **flashing back** to Alice just sitting there" (McNeill, 1985, pp. 361). The gesture anticipated its semantically related word, "there", and terminated several words prior to it.

Similarly, Kita, Alibali and Chu (2017, Fig. 5) discussed a case in which a child produced iconic gestures whose content foreshadowed upcoming speech in his explanation of a Piagetian conservation task. For instance, the content of one of two identical tall glasses of sand was poured into a shallow wide dish. A boy indicated that the tall glass now held more sand than the shallow dish. When asked to explain the reason for this, he said "the bowl is wider, it needs to fill out". As he said, "wider", he moved his hand in a claw shape over the wider dish, and spread his fingers and then moved them back to a claw shape, depicting the concept of spreading before the semantically related words "to fill out".

Quantitative analyses of spontaneous speech have also indicated that speech-gesture synchronisation is not always tight. When adult German speakers were asked to narrate events including manner and path of motion, high proportions of manner and path gestures did not synchronise with their semantically related words, i.e. neither with the manner verb nor the path particle. For example, 55% of iconic path gestures were initiated after the offset of the manner verb and terminated before the path particle. Instead, the gestures were placed in between the two constituents of the motion event which is possible given German main clauses (e.g., "Die Tomate rollt, wie im Video gesehen, den Hügel hinunter." Literal translation: "The tomato rolls, as seen in the video, the hill down"; Here the onset of the stroke occurred after the manner verb "roll" and terminated prior to the onset of the path particle "down") (Fritz, 2018).

Similarly, Morrel-Samuels and Krauss (1992) reported that out of 60 gestures in English narratives, four gestures ended before the semantically related words started. This result might have even underestimated how often gestures preceded the words semantically related. In their study, a group of naive participants watched a video clip of speakers producing speech and gesture, and indicated words that they judged to be related to the gesture; these words were considered to be the semantic affiliates of the gestures. In such a procedure, words that are synchronised with a gesture may be judged to be semantically related to the gesture simply due to the synchrony (temporal relatedness), which would lead to an underestimation of how often gestures do *not* synchronise with semantic affiliates. Taken these studies together, there is clear evidence that gestures are sometimes produced and terminated before their semantic affiliates in speech.

Given that a co-speech gesture sometimes precedes and terminates before its semantically related words in speech, it is important to investigate whether a recipient can form an idea unit of gesture and speech without gesture-speech synchrony. Iconic gestures in isolation are often ambiguous. Although they depict objects, object features, actions or motions, their interpretation is guided by speech and often even depends on speech (Geoffrey Beattie & Shovelton, 1999; Feyereisen, Vandewiele, & Dubois, 1988; Hadar & Pinchas-Zamir, 2004; Krauss, Morrel-Samuels, & Colasante, 1991). Thus, if a recipient is unable to interpret a gesture due to its ambiguity, such a gesture might not be integrated into a discourse model. Previous studies have suggested that the synchrony of a gesture with its speech affiliate is a crucial factor determining whether or not a gesture becomes part of a recipient's discourse model (see Özyürek, 2014 for a review).

The impact of synchrony on the semantic integration of speech and gesture

Investigating the influence of gesture-speech synchrony on integration processes, ERP studies have shown that the synchronisation of semantically affiliated words and gestures is crucial for semantic integration. For instance, Obermeier, Holle, & Gunter (2011) used the homonym disambiguation paradigm described above (Holle & Gunter, 2007), manipulating the overlap of gestures with the semantically related words (homonyms) and task. When a gesture preceded and thus did not temporally overlap with the homonym (e.g., a throwing or dancing gesture appearing at the beginning of the sentence *Sie beherrschte den Ball, was sich im Spiel/Tanz…* "She controlled the ball [= ambiguous noun], which during the game/dance…"), a mismatching

gesture (e.g., throwing versus a dancing gesture) elicited an N400 effect on the target word (e.g., *Tanz* "dance") downstream the sentence only when participants were asked to judge whether the gesture and speech were compatible. The N400 was not present during a semantically "shallow" task, where participants were asked whether a particular movement or word had been present in the preceding trial and therefore did not need to process the gestures semantically. These results confirmed that when speech-gesture synchrony is lost, a gesture does not automatically become part of a recipient's discourse model.

Habets et al. (2011) who manipulated the degree of synchrony between iconic gestures and corresponding isolated verbs in a task that did not explicitly ask participants to semantically process the gestures suggest that a gesture's ambiguity combined with a lack of semantic affiliate serve as reasons for the gesture not being integrated into a person's discourse model. Like Obermeier et al. (2011), Habets and colleagues found that when speech and gesture did not temporally overlap, the gesture was not integrated with the semantic affiliate. This was evident in the absence of an N400 effect on the verb, when comparing matching and mismatching gestures. Habets et al. (2011) hypothesised that when synchrony is lost, recipients interpret the gestures before processing the verb. And because gestures are generally very ambiguous, the interpretation of the gestures is likely to be different from the interpretation of the verbs that are subsequently presented (even in the matched condition). This explanation was supported by the results of a pre-test that investigated whether the gestures presented in their study were semantically transparent without the corresponding speech element. Only in 11% of the cases did the participants' gesture interpretation match the meaning of the verb used in the experiment. Thus, for semantic integration of an isolated word and an iconic gesture, synchronisation of the word and the gesture appears to be crucial.

The impact of gesture ambiguity on gesture-speech integration becomes apparent when we compare Habets et al.'s (2011) findings with findings from a study by Kelly et al. (2004). In the latter study, two objects were placed on a table (i.e., a *short wide dish* and a *tall thin glass*). A person verbally described one of the two objects after providing a gesture that either fitted the description or not. When the gesture was incongruous with speech (e.g., a gesture indicating *shortness* was followed by the word, "tall") they observed that the N400 on the speech was larger compared to when gesture and speech were congruent (e.g., both expressing *shortness*). In both Habets et al.'s (2011) and Kelly et al.'s (2004) studies, gestures were presented before speech. However, participants in Kelly et al.'s (2004) study were able to infer the gestures' meaning from the context, namely the fact that the objects were visually presented together with the gestures. In contrast, in Habets et al.'s (2011) study no disambiguating context was presented. Kelly et al. (2004) found a mismatch effect, whilst Habets et al. (2011) did not.

If speech-gesture synchrony is crucial, are ambiguous gestures that are not synchronised with their semantic affiliate ever integrated into the discourse model? As discussed above, gestures in natural conversations are often not synchronous with the semantically related words. Furthermore, it is unusual that real-world contexts fully disambiguate the gestures (i.e., the situation as in Kelly et al. (2004) is unusual). One potential way to mitigate such a situation is to consider the possibility that discourse information preceding a gesture may constrain the interpretation of the gesture. The language processing literature has shown that discourse information can constrain the possible meaning of upcoming words (see Huettig, 2015 for a review). One source of evidence for such an effect comes from studies using the visual world paradigm (see Huettig, Rommers & Meyer, 2011 for a critical review). In a seminal study, Altmann & Kamide (1999) presented participants with a scene that included an agent (e.g., a boy) together with a number of objects, such as a train set, a cake, a toy car, and a balloon. Crucially, a cake was the only edible object. While viewing this scene, participants listened to one of the two types of sentences, "The boy will eat the cake", or "The boy will move the cake". When presented with a sentence with a constraining verb such as "eat", participants looked towards the target

object (the cake) earlier than when presented with a sentence with a generic verb like "move". Thus, based on the linguistic information received, people anticipate upcoming information. Such anticipatory processing may constrain the meaning of an ambiguous gesture. This constrained meaning of the gesture may be fully integrated into the recipient's discourse model, especially once the gesture's semantic affiliate downstream completely disambiguates the gesture's meaning.

Indirect evidence for this idea comes from a follow-up study of Obermeier et al (2011) where the same homonym stimuli were used but the gesture preceded the homonym (the semantic affiliate) and synchronised with a preceding verb (e.g., "controlled"). This verb did not have any direct semantic relation to the gesture, but Obermeier & Gunter (2015) found an N400 mismatch effect on the disambiguating word later in the sentence (without asking participants to explicitly interpret the gesture). This was interpreted as evidence that gesture can be successfully integrated into a discourse model despite the loss of gesture-affiliate synchrony. Possibly, this successful integration of asynchronous gestures was due to the discourse information preceding the gestures. However, Obermeier & Gunter's study (2015) was not designed to test the influence of preceding discourse on gesture integration, thus we can only speculate about its role in their study.

ERPs reflecting different kinds of speech-gesture integration processes

When an ambiguous gesture precedes and terminates before its semantic affiliate in speech, and the meaning of the gesture becomes gradually clear in discourse, what ERP effects would reflect the speech-gesture integration processes during a task that does not explicitly ask participants to focus on the meaning of the gesture? In such situations, N400 effects may not reflect the integration process at the time when the gesture is seen because the gesture is not initially fully interpretable. In previous studies of speech-gesture integration that elicited the N400 effect at the gesture onset, gestures co-occurred with semantically related words and thus were fully interpretable given the speech context (e.g., Holle & Gunter, 2007; Özyürek et al., 2007). Instead, when the meaning of a gesture is constrained (but not fully disambiguated) by the preceding discourse, we may observe a modulation of anterior ERPs. Semantic ambiguities lead to processing loads reflected in sustained anterior negativities. This is the case for all kind of linguistic ambiguities, including noun-noun homographs (e.g., *organ*; e.g., Hagoort & Brown, 1994), noun-verb homographs (e.g., *the/to watch*; e.g., Federmeier, Segal, Lombrozo, Kutas, M., 2000; Lee & Federmeier, 2009), and referential ambiguities (e.g., van Berkum, 2009). Anterior negativities have also been found during the processing of visual ambiguities such as ambiguous objects (e.g., Dyck & Brodeut, 2015). In all these cases, more ambiguous stimuli lead to more negative ERPs at anterior electrode sites. Importantly, context modulates anterior negativity effects. For linguistic stimuli, semantic constraints caused by preceding linguistic context eliminate frontal negativities of noun–verb ambiguities (e.g., *I knew the meat needed more flavor, but found that it wasn't all that easy to season* versus *I knew the girl threatened more teammates, but commented that it wasn't all that willing to season*; Lee & Federmeier, 2009). Also, presenting ambiguous objects in congruent scenes compared to neutral scenes reduces anterior negativity to the level of unambiguous objects (Dyck & Brodeut, 2015). Similar to this effect of context on ambiguous words and objects, preceding context that constrains an ambiguous gesture's meaning might lead to less negative anterior ERPs.

When the recipient hears the semantically related word of the gesture and the meaning of the gesture becomes fully specified, what ERP effect would reflect this integration process? One possibility is a P600 effect, which is a positive deflection reaching its peak at around 600 ms after the presentation of a target stimulus. This effect tends to be rather long-lasting and comes in the form of a broad peak-less shift (Kutas et al., 2007) and a posterior distribution in experiments with incongruous sentence endings (Van Petten & Luka, 2012). Although this component is traditionally associated with syntactic processing, more

recent research suggests that the P600, more generally, reflects how easily a stimulus can be integrated into the existing mental representation/discourse model (Brouwer, Crocker, Venhuizen, & Hoeks, 2017; Brouwer et al., 2012; Brouwer & Hoeks, 2013). A conflict between new information and the existing discourse model has been found to elicit a larger P600 regardless of whether this conflict arose from syntactic anomalies or semantic anomalies (see Brouwer et al., 2012 for a review). Thus, when a gesture's meaning mismatches the meaning of its semantic affiliate, the resulting conflict leads to a restructuring of the discourse model which elicits a larger P600 compared to gestures that are congruent with the previous/concurrent discourse context (e.g., Cornejo et al., 2009 looking at metaphorical gestures).

The present study

The current study challenges the conclusion that gestures can be integrated into speech only when the gesture synchronises with its **semantic affiliate**. We define "semantic affiliate" of a gesture as the portion(s) of speech near the gesture that is (are) focal points of the speech-gesture semantic link. At a focal point, the semantic overlap between speech and gesture is substantially larger than other portions of speech near the gesture. The term "semantic affiliate" is similar to the term "conceptual affiliate" in de Ruiter (2000), which refers to all portions of speech that are related to the concept expressed in gesture, including portions expressing "high level discourse information" (p.291). However, we argue that it is important to distinguish two types of speech-gesture semantic relationships that are conflated in the concept "conceptual affiliate": (1) discourse contexts that loosely constrain the gesture's meaning versus (2) speech segments that are the focal point(s) of the speech-gesture semantic link, in which semantic overlap between speech and gesture is substantially higher than loosely constraining discourse context. We did not use the term "lexical affiliate" (introduced by Schegloff, 1984) because it assumes that there is a one-word-one-gesture mapping which is often not the case, as pointed out by de Ruiter (2000).

A behavioural experiment (Experiment 1) tested whether a preceding verbal discourse can a) constrain a recipient's interpretation of a gesture when it does not synchronise with its semantic affiliate (starts and terminates before) and b) enable the gesture to be integrated into a recipient's discourse model. In this experiment, participants were explicitly asked to interpret the gestures. Furthermore, an ERP experiment (Experiment 2) investigated the nature of speech-gesture integration processes at two points, namely when recipients encounter a non-synchronous gesture and when they encounter its semantic affiliate, occurring downstream in discourse. In this experiment, participants were not explicitly asked to interpret the gestures.

In both experiments, we used essentially the same set of sentences in which we manipulated the discourse information that preceded the gesture. Gestures depicted actions (e.g., a *picking* gesture) and their semantic affiliates were always verbs (e.g., the verb *picking*). Importantly, gestures never co-occurred with their semantic affiliates, but on a content word (as opposed to a function word; see Obermeier & Gunter, 2015) earlier in the sentence. For instance, the *picking* gesture synchronised with the word *uncle* in the sentence: "I saw that my **uncle** and his two children were *picking* strawberries". As apparent from this example, the content word synchronous with the gesture could not disambiguate the gesture.

In particular, we manipulated the semantic relation between the gesture and its preceding discourse by using two types of introductory sentences (semantically related versus semantically unrelated to the gesture). In the Unrelated Discourse Condition, the introductory sentence did not provide any information to constrain the interpretation of the gesture (e.g., "At the beginning of the week the weather was dreadful" for a following *picking* gesture). In contrast, in the Related Discourse Condition, the introductory sentence provided the recipient with information that was related to the gesture's meaning ("Some of

the strawberries in the garden were already ripe" for the following *picking* gesture). The related introductory sentence could thus constrain the possible interpretations of the gesture.

Experiment 1 – Behavioural experiment

The behavioural experiment tested whether our discourse manipulation had an effect on how participants interpret the meaning of iconic gestures. Previous studies found that iconic gestures are difficult to interpret without accompanying speech (cf. Geoffrey Beattie & Shovelton, 1999; Feyereisen et al., 1988; Habets et al., 2011; Hadar & Pinchas-Zamir, 2004; Holle & Gunter, 2007). And even when the gestures are presented in combination with co-occurring speech (i.e., a co-occurring phrase), people often cannot choose the correct meaning of gestures in a forced-choice paradigm (Hadar & Pinchas-Zamir, 2004). In our behavioural experiment, participants heard and watched videos up to, but not including the gesture's semantic affiliate, and then described their interpretation of the gesture. The semantic affiliate was not presented so that it would not give away the intended meaning of the gesture. Participants were asked to interpret gestures within a sentence context and to provide some information about how they interpreted the gestures. If the preceding discourse indeed affects gesture interpretation, the number of different gesture interpretations that different participants come up with should be lower in a Related Discourse Condition compared to an Unrelated Discourse Condition. Also, gesture interpretation should be easier in the Related Discourse Condition than in the Unrelated Discourse Condition. Furthermore, the behavioural experiment validated the gesture match/mismatch manipulation of our stimuli for the ERP experiment (Experiment 2). If our manipulations worked as intended, then participants' gesture interpretations (e.g., of the *picking* gesture) should match the semantic affiliate in the Match Condition (e.g., the verb *picking*) more closely than in the Mismatch Condition (e.g., the verb *watering*).

Methods

Participants

Seventeen native speakers of English took part in the behavioural experiment. One participant was excluded from the analysis due to severe visual impairment. The remaining sixteen participants (mean age = 19 years, SD = 0.8, all female) all studied Psychology at the University of Birmingham and received course credits for their participation.

Material

We created 53 stimuli, each consisting of an introductory sentence, a target sentence and a gesture. The gestures and their semantic affiliates (verbs) were part of the target sentences. Discourse information was manipulated by varying whether the introductory sentences provided helpful information for the interpretation of the gesture or not. In the Related Discourse Condition the introductory sentence provided information that was related to the gesture's meaning and could therefore help to interpret it, while in the Unrelated Discourse Condition this was not the case (see Table 1 for example stimuli). In both conditions, the gesture was placed on a content word (2–3 syllables long) towards the beginning of the target sentence. This word could not disambiguate the gesture without discourse information (e.g., a *watering gesture* synchronised with the word "uncle"). The semantic affiliate was always the verb of the target sentence (e.g., watering, picking), which was linguistically encoded further downstream in the sentence (e.g., I saw that my uncle and his two children were *picking* strawberries, with a gesture co-occurred with "uncle").

A mismatch paradigm was used. In the Gesture Match Condition, the gesture matched the target-verb (e.g., speech: "picking"; gesture: *picking*). In the Gesture Mismatch Condition the gesture did not match the target-verb (e.g., speech: "picking"; gesture: *watering*). Importantly, up until the verb, the semantic fit of the Match Gesture and the Mismatch

Table 1

Example stimulus for Experiments 1 and 2 for all four stimulus sets. The onset of the first word in bold in the target sentence co-occurred with the onset of the gesture. For the behavioural experiment (Experiment 1), stimuli were presented without the target verb (the second word in bold) and any following words. For the ERP study (Experiment 2), full sentences were presented and ERPs were time-locked to the onset of both words in bold in the target sentence. The words underlined in the introductory sentences provided information related to the meaning of the gesture in the target sentence. The four stimulus sets were counter-balanced across participants. Target verbs and speech elements between the gesture and the target verb were counter-balanced as follows: Stimulus Sets 1 and 2 differed in terms of the elements between the gesture and the target verb. The same applied to Stimulus Sets 3 and 4. The target verbs were the same in Stimulus Sets 1 and 2 but different from those in Stimulus Sets 3 and 4.

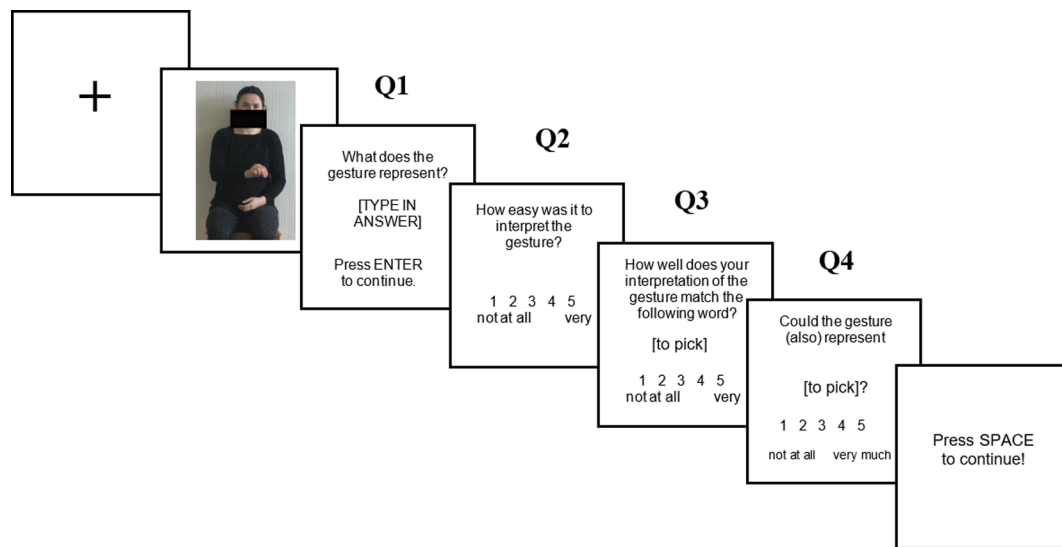| Condition	Introductory Sentence	Target Sentence	Gesture (match/mismatch)
Stimulus Set 1			
Unrelated Discourse Condition	At the beginning of the week the weather was dreadful.	I saw that my **uncle** and his lovely wife were **picking** strawberries.	Match ("picking gesture")
	At the beginning of the week the weather was dreadful.	I saw that my **uncle** and his lovely wife were **picking** strawberries.	Mismatch ("watering gesture")
Related Discourse Condition	Some of the <u>strawberries</u> in the garden were already <u>ripe.</u>	I saw that my **uncle** and his two children were **picking** strawberries.	Match ("picking gesture")
	Some of the <u>strawberries</u> in the garden were already <u>ripe.</u>	I saw that my **uncle** and his two children were **picking** strawberries.	Mismatch ("watering gesture")
Stimulus Set 2			
Unrelated Discourse Condition	At the beginning of the week the weather was dreadful.	I saw that my **uncle** and his two children were **picking** strawberries.	Match ("picking gesture")
	At the beginning of the week the weather was dreadful.	I saw that my **uncle** and his two children were **picking** strawberries.	Mismatch ("watering gesture")
Related Discourse Condition	Some of the <u>strawberries</u> in the garden were already <u>ripe.</u>	I saw that my **uncle** and his lovely wife were **picking** strawberries.	Match ("picking gesture")
	Some of the <u>strawberries</u> in the garden were already <u>ripe.</u>	I saw that my **uncle** and his lovely wife were **picking** strawberries.	Mismatch ("watering gesture")
Stimulus Set 3			
Unrelated Discourse Condition	At the beginning of the week the weather was dreadful.	I saw that my **uncle** and his lovely wife were **watering** the strawberries.	Match ("watering gesture")
	At the beginning of the week the weather was dreadful.	I saw that my **uncle** and his lovely wife were **watering** the strawberries.	Mismatch ("picking gesture")
Related Discourse Condition	Some of the <u>strawberries</u> in the garden were already <u>ripe.</u>	I saw that my **uncle** and his two children were **watering** the strawberries.	Match ("watering gesture")
	Some of the <u>strawberries</u> in the garden were already <u>ripe.</u>	I saw that my **uncle** and his two children were **watering** the strawberries.	Mismatch ("picking gesture")
Stimulus Set 4			
Unrelated Discourse Condition	At the beginning of the week the weather was dreadful.	I saw that my **uncle** and his two children were **watering** the strawberries.	Match ("watering gesture")
	At the beginning of the week the weather was dreadful.	I saw that my **uncle** and his two children were **watering** the strawberries.	Mismatch ("picking gesture")
Related Discourse Condition	Some of the <u>strawberries</u> in the garden were already <u>ripe.</u>	I saw that my **uncle** and his lovely wife were **watering** the strawberries.	Match ("watering gesture")
	Some of the <u>strawberries</u> in the garden were already <u>ripe.</u>	I saw that my **uncle** and his lovely wife were **watering** the strawberries.	Mismatch ("picking gesture")

Fig. 1. Example trial used in the behavioural study (Experiment 1). In this example, the target verb was "to pick", which was not presented in the video stimulus in Experiment 1.

Gesture did not differ in either the Related Discourse Condition or the Unrelated Discourse Condition. To ensure that there was no temporal overlap between the gesture and the verb, we separated them by inserting a number of words with in total 5–7 syllables (number of words varied). The meanings of the words placed between gesture and verb were kept neutral in relation to the verb's and the gesture's meaning. This was to prevent the participants from predicting the critical verb and to avoid providing more information that could disambiguate the gesture.

The structure of the target sentences was very similar across stimuli. They all started with one of the following phrases: "I could see/hear", "I saw", "I noticed", "I was told". Despite the target verb occurring towards the end of the sentence, it never appeared in the sentence final position. This was to avoid sentence-final wrap-up effects in the ERP study (i.e., Experiment 2), which increase processing cost due to global processing of the sentence (Hagoort, 2003; Osterhout, 1997).

Each participant was presented with the introductory sentence from the Related Discourse Condition and the Unrelated Discourse Condition twice, once with a matching and once with a mismatching gesture; thus, she/he was presented with the target sentence four times (i.e., matching and mismatching gesture in both discourse conditions). Hearing the target sentence four times throughout the experiment might have resulted in the participants memorising parts of the sentences which could help them with predicting the verb. We therefore changed the words between the gesture and the verb across conditions, keeping the total number of syllables the same (see Table 1 for an example stimulus).

The words between the gesture and the verb were counter-balanced across participants. Furthermore, we counter-balanced across participants which target-verb of the match/mismatch combination the participants had to respond to (e.g., half of the participants were presented with "watering", the other half with "picking") (see Fig. 1 for an example trial). In sum, this led to four different stimulus sets (see Table 1) of which each participant was presented with only one.

Stimulus recordings & editing

Speech was recorded in a soundproof booth by a female native speaker of English. Introductory sentences and target sentences were recorded separately and were later combined using Adobe Premiere Pro (www.adobe.com). This allowed us to counterbalance the target-verb and the speech elements between the gesture and the target verb (see counterbalancing in Table 1).

For the gestures, we video-recorded a female person performing the gestures (with 29 frames per second). Although a few verbs occurred twice in the stimuli for a given participant, we used different gestures that fitted the discourse context (e.g., to pick strawberries vs. to pick an apple). Thus, even when a verb was repeated, we did not repeat the same gesture with the same meaning. To make the gesture videos appear natural, the gesture actor was instructed to utter the beginning of the target sentence and to produce the gesture while uttering the word with which the gesture would later synchronise. After executing the gesture, her hands returned to the resting position (i.e., her lap) while a few more seconds were recorded of her sitting still. Her speech was not used in the stimuli.

Then, we temporally aligned separately recorded gesture and speech, using the video-editing software: gesture stroke onset was synchronised with the onset of the target noun. On average, strokes lasted for 922 ms (SD = 364 ms). Gesture onset occurred on average 5.8 sec (SD = 0.7) after the onset of the introductory sentence. Onset of the target-verb occurred on average 8.1 sec (SD = 0.9) after the onset of the introductory sentence, leaving on average 2.2 sec between gesture onset and verb onset. Importantly, the gesture's stroke never synchronised with the verb but always terminated prior to the verb's onset (on average 1288 ms before verb onset). Finally, the mouth of the actress was covered with a black box because her mouthing did not match the audio track. In total 848 stimuli were created.

Procedure

Participants were randomly assigned to one of the four stimulus sets. The participants' task was to listen and watch each video carefully and afterwards answer four questions about the gesture they had seen in the video (see Fig. 1 for study design). Videos were presented up to the word immediately *before* the target verb so that initial gesture interpretations were not affected by the verb. Question 1 was an open question, asking the participants to type in what they thought the gesture represented. Participants were told that their answer could range from one to four words. For Questions 2 to 4, participants were asked to provide ratings on a scale from 1 (not at all) to 5 (very) (see Fig. 1 for the wording of the questions). Question 2 asked participants how difficult it was to interpret the gesture. Question 3 asked them to rate how similar their own interpretation of the gesture (response to Question 1) was to a specific target-verb. The target-verb either matched or mismatched the gesture. Which of the two target-verbs (e.g., "picking" or "watering") the participants saw depended on the stimulus set they were assigned to (see Table 1). Question 4 asked participants to rate how well the target-verb
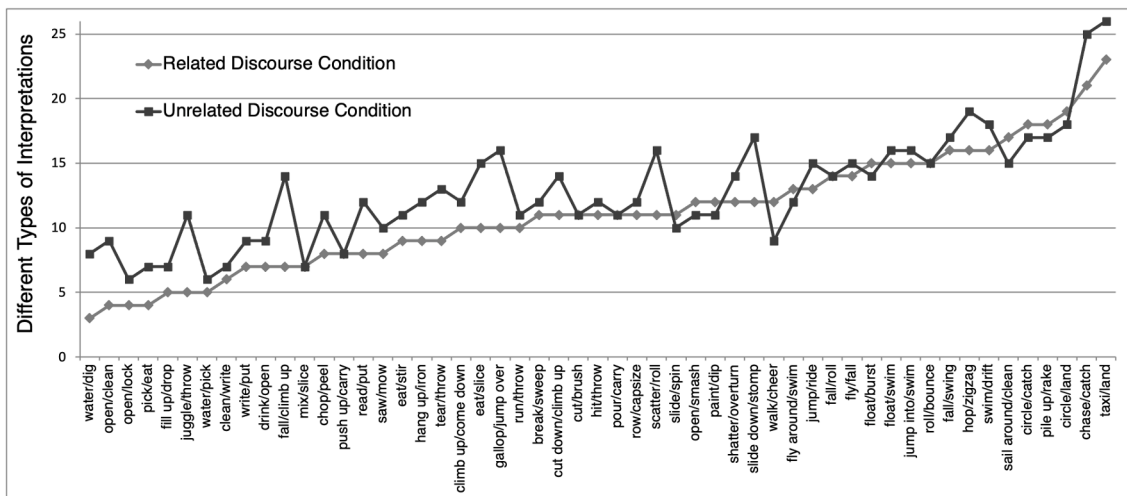
Fig. 2. Diversity of responses across participants to the question: "What does the gesture represent?", split into Related and Unrelated Discourse conditions. A higher value indicates that the participants provided a larger set of distinct interpretations for a gesture. Items in the X-axis are ordered from the lowest number of different types of gesture interpretations to the highest number of different types of gesture interpretations.
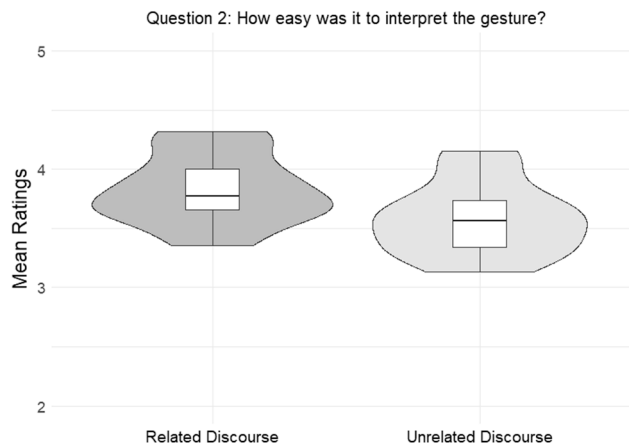

Fig. 3. Ratings for how easy it was to interpret the gesture, for the Related and Unrelated Discourse Condition (max score (easiest) = 5), as probed by Question 2 in Experiment 1. Violin plots show the kernel probability densities of the data at different values. The horizontal line inside each box plot denotes the median of the data. Box limits indicate the interquartile range. Tails of the violins were trimmed to the range of the data.

fit the gesture regardless of their own interpretation of the gesture. Thus, Question 4 asked the participant to reanalyse the gesture's meaning.

Participants were presented with each verb in its infinitive form (e.g., to swim) to avoid any confusion with nouns (e.g., to iron vs. the iron). The study was self-paced and the participants could take breaks whenever needed. In total, each participant responded to 212 stimuli, presented in four blocks. The order of the blocks was counter-balanced and the trials within the blocks randomised. The experiment took approximately one hour to complete.

Results

All the data files used in this study are available at the following open OSF archive: https://osf.io/wr7a5/. Question 1 had asked participants to write down their own interpretation of the gesture. We conducted a by-item analysis of the responses. We determined response diversity by counting the total number of differing responses for a particular gesture across all participants (see Fig. 2). Responses with the same verb in different grammatical forms were collapsed (e.g., fly, flying, to fly were all counted as 'to fly' responses). So were combinations of the verb with other words (e.g. bird flying, flying around). Answers without a verb (e.

g., down, wide, above) were counted as separate responses. Participants were allowed to use up to four words in a response, and if a response included more than two interpretations (e.g., "bird flying somebody jumping"), both interpretations were included in the analysis. This measure thus provides a type count of differing interpretations to each gesture. We found that the diversity of responses per item was significantly larger in the Unrelated Discourse Condition compared to the Related Discourse Condition ($t(52)$, 5.56, $p < .001$).

For the second question, which asked participants how difficult it was to interpret the gesture, a significant effect of Discourse type was found ($t(15) = 4.93$, $p < .001$), with gestures shown in the Related Discourse Condition reported to be easier to interpret than those shown in the Unrelated Discourse Condition (see Fig. 3).

For Questions 3 and 4, 2.5% of the responses had to be excluded from the analysis due to a technical error that led to a wrong verb being presented in four stimuli. Results for Question 3, which asked how well the participant's own interpretation of a gesture matched the target word, are summarised in Fig. 4. Ratings were analysed with a repeated measures ANOVA with Gesture Match (Match/Mismatch) and Discourse (Related Discourse/ Unrelated Discourse) as within-subject independent variables. A significant interaction between Discourse and Gesture
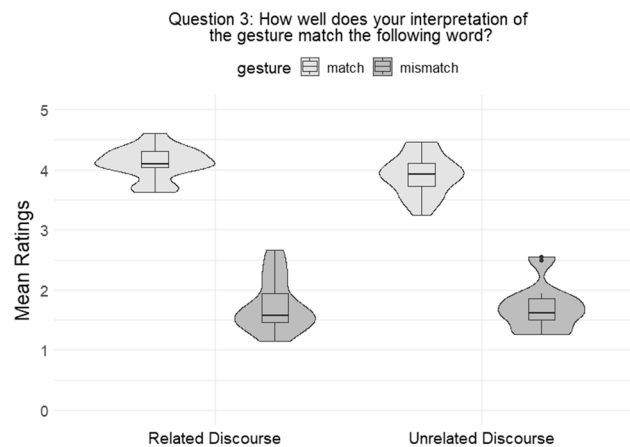
Fig. 4. Effects of Gesture Match and Discourse Relatedness on mean ratings for how well the gesture matched the target words (i.e., verbs), as probed by Question 3, max score (best match = 5) in Experiment 1. Violin plots show the kernel probability densities of the data at different values. The horizontal line inside each box plot denotes the median of the data. Box limits indicate the interquartile range. Tails of the violins were trimmed to the range of the data.

Match was found ($F(1,15) = 5.38$, $p = .035$), as well as significant effects of Gesture Match ($F(1,15) = 654.0$, $p < .001$) and Discourse ($F(1,15) = 8.1$, $p = .012$). Post-hoc paired t-tests revealed significantly higher ratings for matching gestures compared to mismatching gestures for both Related ($t(15) = 21.3$, $p < .001$) and Unrelated Discourses ($t(15) = 26.7$, $p < .001$), confirming our gesture-verb matching manipulation. Importantly, matching gestures were rated to be better matches with the verb in the Related Discourse condition than in the Unrelated Discourse conditions ($t(15) = 3.2$, $p = .006$). This was not the case for mismatching gestures ($t(15) = .4$, $p = .667$). Thus, a related discourse guided the interpretation of the gesture the way we intended, while the preceding discourse did not affect the match of a mismatching gesture with the verb.

Finally, Question 4 asked how well the target-verb fitted the gesture regardless of their interpretation of the gesture. Here we did not find an effect of Discourse ($F(1,15) = 0.1$, $p = .797$), nor a significant interaction between Gesture Match and Discourse ($F(1,15) = 3.3$, $p = .091$). Only Gesture Match yielded a significant effect ($F(1,15) = 340.7$, $p < .001$), with matching gestures being rated to be a better representation of the verbs than mismatching gestures. These results again confirmed our manipulation of gesture match/mismatch with the verb. Results are summarised in Fig. 5.

Discussion experiment 1 - behavioural experiment

Experiment 1 tested whether a preceding discourse that is semantically related to a gesture's meaning constrains gesture interpretation when the gesture does not synchronise with its semantic affiliate. Three results showed that the participants' interpretations of gestures were influenced by our discourse manipulation. First, the presence of a preceding discourse that was related to the gesture's meaning reduced response diversity in participants' interpretations of the gesture (Question 1). In other words, participants agreed with each other in gesture interpretation more often in the Related Discourse Condition than in the Unrelated Discourse Condition. Second, participants rated gestures that were presented in the Related Discourse Condition as being less difficult to interpret than in the Unrelated Discourse Condition (Question 2). Third, participants rated their gesture interpretations to be more similar to our matching target verbs in the Related Discourse Condition than in the Unrelated Discourse Condition. In contrast, for the mismatching gestures, we did not find a difference between the two discourse conditions (Question 3). Importantly, while the discourse affected participants' gesture interpretations, the gestures represented the matching target verbs well: Participants rated the gestures as very good matches with the target verbs independently of their own interpretation, and this gesture reanalysis was not affected by discourse (Question 4).

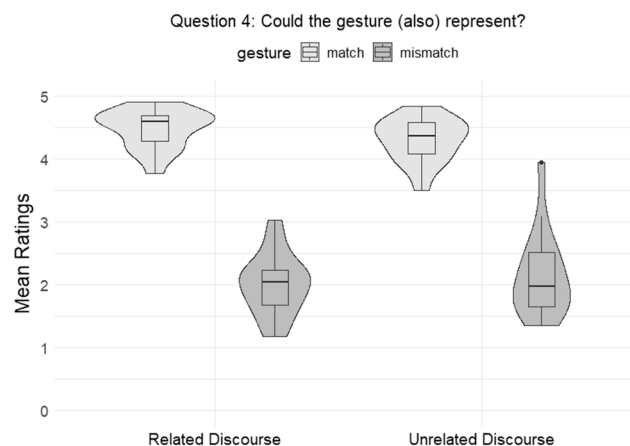What do these results tell us about the role of preceding discourse

Fig. 5. Ratings of how well participants' interpretation of gestures match the target words (verbs) in the Related and Unrelated Discourse Condition, as probed by Question 4, max score (best match) = 5 in Experiment 1. Violin plots show the kernel probability densities of the data at different values. The horizontal line inside each box plot denotes the median of the data. Box limits indicate the interquartile range. Tails of the violins were trimmed to the range of the data.

information in gesture interpretation? They show that, similarly to speech processing (cf. van Berkum, 2008), the meaning of a gesture is constructed in a discourse-dependent way. More precisely, meaning construction does not depend exclusively on words concurrent with the gesture, but also depends on preceding discourse information. This does not mean that discourse cues fully disambiguate a gesture's meaning. Instead, they might constrain the number of possible interpretations, as we have seen in the lower response diversity in our Related Discourse Condition.

Many of the previous behavioural studies that examined the role of speech on gesture *interpretation* did not examine gestures in a discourse context. They compared gesture interpretations when gestures were presented without speech and with co-occurring speech elements (i.e., a phrase or clause) (Geoffrey Beattie & Shovelton, 1999; G. Beattie & Shovelton, 2002; Hadar & Pinchas-Zamir, 2004). Or they presented gestures without any speech (Feyereisen et al., 1988; as part of a pretest: Habets et al., 2011). Since these studies took the gestures out of their discourse context (i.e., a narrative), they could not provide a full picture of how people interpret iconic gestures in a task where they are explicitly asked to do so. Our study therefore adds to the literature, highlighting that gesture interpretation is discourse-dependent.

Our results on the influence of discourse information on gesture interpretation are also in line with Özyürek et al.'s (2007) ERP finding that gesture processing is influenced by the information provided in the preceding discourse. Özyürek et al.'s (2007) investigated cases where the gesture is synchronised with its semantic affiliate, and the preceding discourse fully disambiguated the meaning of gesture (or fully made clear that the gesture does semantically fit with the preceding discourse). In contrast, we investigated how preceding context loosely constrains the interpretation of gestures that are not synchronised with their semantic affiliate.

Experiment 2 – ERP experiment

Experiment 2 used ERPs to examine whether the information in a preceding discourse that is loosely related to a gesture's meaning can influence processing of the gesture, and whether the information encoded in such a gesture can influence processing of its semantic affiliate later in the sentence. We used the same materials and design as in Experiment 1, but presented stimuli without explicitly asking participants to interpret the gestures.

ERPs time-locked to the gesture's onset

In order to investigate the effect of discourse on gesture interpretation we compared ERPs time-locked to the gesture's onset in the Related Discourse Condition with those in the Unrelated Discourse Condition. In Experiment 1, we saw that gesture interpretation was more constrained (but not fully disambiguated) by preceding discourse information in the Related Discourse Condition than in the Unrelated Discourse Condition. Moreover, participants rated the gesture interpretation task as being more difficult in the Unrelated Discourse Condition. Since the gesture was placed on a word that does not help to disambiguate it, participants seem to have attempted to find information in the preceding discourse that could constrain the meaning of the gesture. Since constraining an ambiguous gesture is somewhat similar to disambiguating a word or object (Lee & Federmeier, 2009; Dyck & Brodeut, 2015), the Unrelated Discourse condition might lead to more negative anterior ERPs when compared to the Related Discourse condition.

In contrast, we did not expect an N400 effect on gestures when comparing the two discourse conditions because the preceding discourse only roughly constrained but did not disambiguate the gestures in our behavioural experiment. In other words, even the preceding discourse in the Related Discourse Condition should not strongly prime the construction of the gesture's meaning. Brouwer and colleagues (Brouwer et al., 2012) proposed that the N400 reflects the retrieval of conceptual knowledge (e.g., retrieval of lexical semantic representations). The equivalent process for co-speech gesture may be the construction of an unambiguous meaning of a gesture. However, in the current experiment, even in the Related Discourse condition, the context was not strong enough to allow the construction of the gesture's meaning.

We did not expect a P600 effect on gestures either, when comparing the two discourse conditions. This was because the P600 is a post-semantic effect that restructures the discourse model. Because the preceding discourse only roughly constrained but did not disambiguate the gestures, post-semantic processes were unlikely triggered to integrate the gesture's meaning into a discourse model.

ERPs time-locked to the gesture's semantic affiliate

In order to investigate whether the information provided by a gesture can influence processing of its semantic affiliate that occurs later in the sentence, we examined ERPs time-locked to the onset of the gesture's semantic affiliate (i.e., the verb). There were two possible outcomes.

First, if synchrony of gesture and its semantic affiliate was crucial for gesture integration when participants are not explicitly asked to interpret the gesture (e.g., Habets et al., 2011; McNeill, 1992), a mismatch between gesture and semantic affiliate (i.e., the verb) should not elicit any effects (neither N400, nor P600) on the semantic affiliate. This should be the case regardless of the nature of the discourse preceding the gestures.

Second, if gestures that are not synchronous with their semantic affiliates *can* be integrated into the discourse model (see Obermeier & Gunter, 2015), then we would expect to see a mismatch P600 effect, but not an N400 effect, on the semantic affiliate downstream in the sentence. If gestures that are not synchronous with their semantic affiliates had been integrated into the discourse model, then the integration of the semantic affiliate might trigger a reanalysis of the gesture's meaning, which up to this point should still be vague, in an attempt to incorporate the updated meaning of the gesture into the discourse model. This process should be more effortful in the mismatch condition than in the match condition, leading to a P600 effect. It was not clear whether the P600 effect would differ according to the discourse preceding the gesture (related vs. unrelated). However, how constraining the preceding discourse is might have an impact on the reanalysis process, and thus on the P600 effect. The P600 effect might be more pronounced when the preceding discourse was related to the gesture than when it was unrelated because, as seen in Experiment 1, a related discourse should more strongly constrain the interpretation of the gesture.

In contrast to the P600 effect, it was unlikely that we would see an N400 mismatch effect on the semantic affiliate (in either discourse condition). Experiment 1 showed that discourse information constrained gesture interpretation, but the interpretation often did not converge on the semantic affiliate (i.e., the action verb in the second sentence). We found that only 47% of the interpretations of the gestures matched our target-verbs in the Related Discourse Condition, and 41% in the Unrelated Discourse Condition. Thus, in many cases gestures were still ambiguous. It was therefore unlikely that they could consistently prime the semantic retrieval of their affiliates and lead to an N400 effect, as found by Wu & Coulson (2010). Note that Obermeier and Gunter (2015) did find a mismatch N400 effect on target words that occurred downstream the sentence when they presented gestures synchronous with content words and a task similar to us. However, their content words may have constrained the meaning of their gestures more strongly than in our study. Similarly, in Wu & Coulson's studies (2005; 2007; 2010), the presented gesture appeared to be unambiguous enough to function as primes for following picture/word probes (Wu & Coulson, 2005; 2007) and to facilitate lexical retrieval of a gesture's semantic affiliate presented synchronously with the gesture (Wu & Coulson, 2010). We deliberately chose content words that could not disambiguate the gestures. Therefore, in our study, disambiguation was only possible at the semantic affiliate downstream the sentence.

Methods

Participants

In total 38 participants took part in Experiment 2. None of the participants took part in Experiment 1. One participant was excluded from the analysis because of being bilingual. Furthermore, five participants were excluded due to excessive artefacts (more than 25% of all trials). Thus, 32 participants were included in the analyses (mean age = 20.1, SD = 3.4, 24 female, 29 right-handed). These were all monolingual English native speakers with normal or corrected-to-normal vision and did not report any hearing impairment. Note that we kept left-handed participants in the analyses to better represent the wider population, which includes approximately 10% left-handers (cf. Willems, van der Haegen, Fisher, & Francks, 2014). All participants gave written informed consent and received either course credits or £7 per hour as compensation.

Material

The material was the same as that of Experiment 1. However, we excluded some of the trials from the original material. First, responses to Question 3 of Experiment 1 ("How well does your interpretation of the gesture match the following word?") showed that three items showed a very small gesture match effect (i.e., the difference between the scores for matching and mismatching gesture) in the Related Discourse condition (<0.5 on the Likert Scale) (e.g., swimming gesture versus floating gestures for the verb "to swim"). Second, another item (the verb pair "scatter/roll") was excluded because its mismatching gesture (i.e., "rolling marbles across the floor") was not perceived as mismatching (rated as 3.625 out of 5). Finally, another item was excluded because the rating for the match gesture was lower than for the mismatch gesture (i. e., the verb pair "catch/chase"). Taken together, forty-eight out of the 53 stimuli used in the behavioural experiment were included in the ERP experiment.

Procedure

Participants were told before the study that it investigated mechanisms underlying sentence comprehension. The experiment was conducted in a sound-proof booth. The videos were presented on a computer screen placed ~ 90 cm away from the participants. Audio was presented via speakers. Participants were instructed to watch and listen carefully. Furthermore, since each trial was very long (2 sentences, approximately 9 s), participants were told that they could blink when they saw the fixation cross and during the first sentence of each trial, but should not blink during the second sentence.

In order to keep participants' attention, a comprehension task was included. To avoid that participants only paying attention to a specific part of the sentence, our comprehension questions either targeted the sentence's subject, the verb (semantic affiliate) or the phrase inserted between the gesture and the semantic affiliate. After approximately five trials (randomised), participants were asked a yes/no comprehension question about the sentence they had just heard. For instance, for the example stimulus presented in Table 1, the question was "Did the speaker's uncle pick strawberries?". In this case, the answer was *yes*. For questions for which the correct answer was *no*, one word was changed from the actual stimulus sentence (e.g., "Did the speaker's **aunt** pick strawberries?"). All questions were about the target sentence and not the introductory sentence.

The stimulus presentation is illustrated in Fig. 6. First, participants saw a fixation cross for 1,000 ms, followed by the stimulus video. If a comprehension question followed the video, it was displayed in the centre of the screen. No time limit was given for responding to the question. Since the behavioural data was for only a subset of our trials (39 out of 192 items) we decided to include all trials into the ERP analysis regardless of whether the participant's response was correct or not (cf. Obermeier & Gunter, 2015).

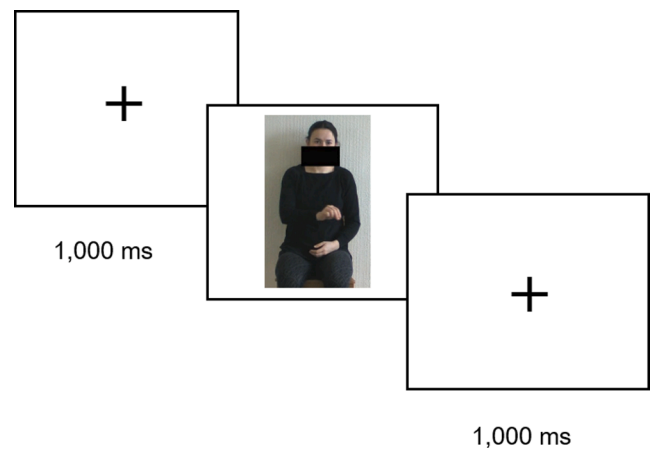Before the actual experiment, participants were presented with five

Fig. 6. Stimulus presentation in Experiment 2 (ERP experiment). Unlike in Experiment 1, the video presented full stimulus sentences as illustrated in Table 1, including the target verb.

practice trials. The purpose of the practice was twofold. First, it enabled the participants to get used to how the stimuli were presented. Second, it enabled the experimenter to check whether the participant understood the instructions on when to blink and when not to blink. The experiment was divided into four blocks, each containing 12 stimuli from each of the four conditions, randomly presented. The order of these blocks was counter-balanced across participants. Taken together, the four different sets of stimuli introduced above and the counter-balancing of the stimulus order resulted in 16 different versions of the experiment. Each block took approximately 9 min to complete. After each block the participant was able to take a break. In total, the experiment lasted about 50 min.

EEG recording and analysis

Electroencephalograms (EEG) were recorded with a 64 BioSemi ActiView EEG system. Electrodes were placed and labelled according to the international 10–10 system. EEG data were sampled at 512 Hz. All electrodes were re-referenced offline to averaged mastoids. Electrooculograms (EOG) monitored vertical and horizontal eye movements. Data were filtered offline using a band-pass filter of 0.1–30 Hz. Additionally, for presentation purposes only, a low pass filter was applied (10 Hz). For each trial, two epochs were created. The first epoch was time-locked to the onset of the gesture; the second epoch was time-locked to the onset of the verb of the second sentence, i.e. the semantic affiliate of the gestures. Epochs started 200 ms pre stimulus onset and lasted 1200 ms. Since the stimulus videos were quite long and despite our efforts to keep the number of blinks on target words to a minimum, our datasets included large proportions of ocular artefacts. We therefore corrected eye-blinks using an Independent Component Analysis (procedure implemented in EEGlab, see Delorme & Makeig, 2004, and following the procedure described in Nunez, Nunez, and Srinivasan, 2016). After eye-blink correction, trials were rejected according to the following criteria: We utilized an automatic artefact rejection procedure with a 200 ms sliding window and 50 ms steps. If any of the electrodes (except EOG) exceeded a $\pm$ 100 μV threshold within an epoch, the epoch was excluded from the analyses. Additionally, all EEG data were inspected manually and any epochs with clear artefacts were excluded.

For the analysis of ERPs time-locked to gesture onset, match and mismatch trials were collapsed because at this point participants could not have been affected yet by the gesture–verb match/mismatch manipulation. Based on the rejection criteria described above, on average, 93.8% of the epochs time-locked to the gesture onset were kept in the analysis.

For the analysis of ERPs time-locked to the verb, 92.7% of all epochs were kept in the analysis.

For the analyses, we followed previous studies on the processing of

metaphorical gestures within a sentence context, covering sites of both N400 and P600 effects (Cornejo et al., 2009; Ibáñez et al., 2010, 2011). This meant we picked five ROIs, with seven electrode sites each: Anterior Left ("AL" = Fp1, AF7, AF3, F7, F5, F3, F1), Anterior Right ("AR" = Fp2, AF8, AF4, F8, F6, F4, F2), Centre (FC1, FCz, FC2, C1, C2, Cz, CPz), Posterior Left ("PL" = P1, P3, P5, P7, PO3, PO7, O1), Posterior Right ("PR" = P2, P4, P6, P8, PO4, PO8, O2). For the analysis time-locked to the gesture, we were not only interested in N400 and P600 effects, but also in an anterior negativity. Previous studies found anterior negativities to be strongest over anterior left and/or anterior right sites. However, since the effect might be located differently for gestures, we picked regions across the whole skull, including Centre Left (FT7, FC5, FC3, T7, C5, C3, CP3) and Centre Right (FT8, FC6, FC4, T8, C6, C4, CP4).

We analysed the ERP data by fitting linear mixed effects models in RStudio (R Core Team, 2018) using the lmer function of the lme4 package (Bates et al., 2015). We examined single trial mean amplitudes in 100 ms windows between 200 ms and 1200 post stimulus onset. By doing so, we covered all time-windows of ERP components that might be associated with semantic processing of the gesture and its semantic affiliate. Statistics were not corrected for multi-comparisons in the ten different time-windows as Bonferroni corrections, for example, would have been too conservative. We instead checked whether any significant effects were robust across time windows and regions (i.e., observable in adjacent windows and regions) since such effects were unlikely to be just false positive.

For the analyses time-locked to gesture onset, fixed effects included Discourse (Related versus Unrelated Discourse Condition) and ROI (7 ROIs) plus their interaction. For each model, we started with a random effect structure including intercepts and random slopes of Discourse and ROI for both Subjects and Items. If this model did not converge, we removed random slopes in the following order: ROI for Subjects, ROI for Items, Discourse for Subjects, Discourse for Items. The factor ROI was coded as dummy variable with Anterior Left as the reference level, because anterior negativities have often been found to be strongest in this area. For the Discourse factor, the reference level was set to Related Discourse Condition. If we found a significant ROI × Discourse interaction, we would fit models with Discourse as fixed effect in each ROI, as posthoc tests.

We applied the same analysis strategy to ERPs time-locked to the verb (i.e., the semantic affiliate of the gesture), but added the fixed factor Gesture Match (Match versus Mismatch) and used the Posterior Right ROI as the reference ROI. The latter was chosen because we expected to find a P600 effect and a P600 should be strong at these sites. The reference levels were Match for the Gesture Match factor, and Related Discourse for the Discourse factor. The random effects structure was the same as for the analyses time-locked to the gesture with the addition of random slope and intercept of Gesture for Subject and Item. If we found a significant Discourse × Gesture Match × ROI interaction, we would fit separate models with the fixed factors Discourse and Gesture Match as well as their interaction in each ROI, as posthoc tests.

Results

All the data files and R scripts used in this study are available at the following OSF archive: https://osf.io/wr7a5/.

Behavioural results

The average response accuracy to the comprehension questions was 94.0% (SD = 5.2), suggesting that participants paid attention to the stimuli.

ERP results

Fig. 7 shows the ERP waveforms time-locked to the gesture onset. While these seem to show more negative ERPs in the Unrelated Discourse condition than the Related Discourse condition at anterior and left electrode sites, these differences were not significant. None of the time windows showed a significant main effect of Discourse or a significant Discourse × ROI interaction (see Table 1 in the Supplementary Materials).

Fig. 8 shows the ERP waveforms for a selection of electrodes time-locked to the target verb (i.e., the semantic affiliate of the gesture). Table 2 shows the results of the analyses. There were no significant
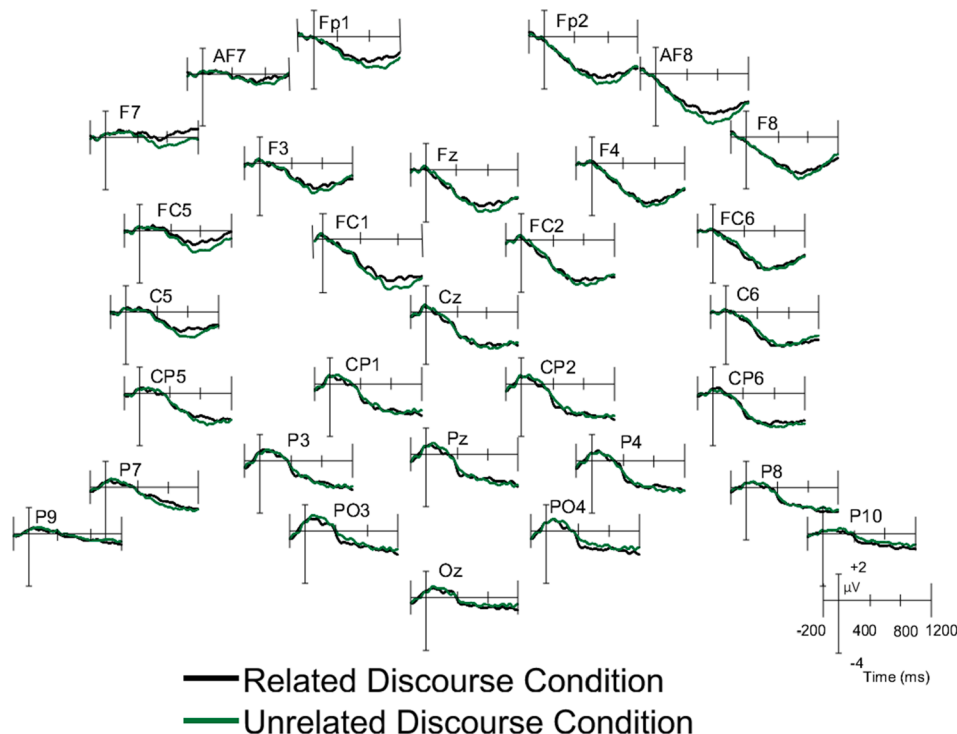

Fig. 7. ERP responses at a subset of electrodes in the Related Discourse Condition compared to the Unrelated Discourse Condition, time-locked to gesture onset. Note negativity is plotted downwards. None of the differences were significant.
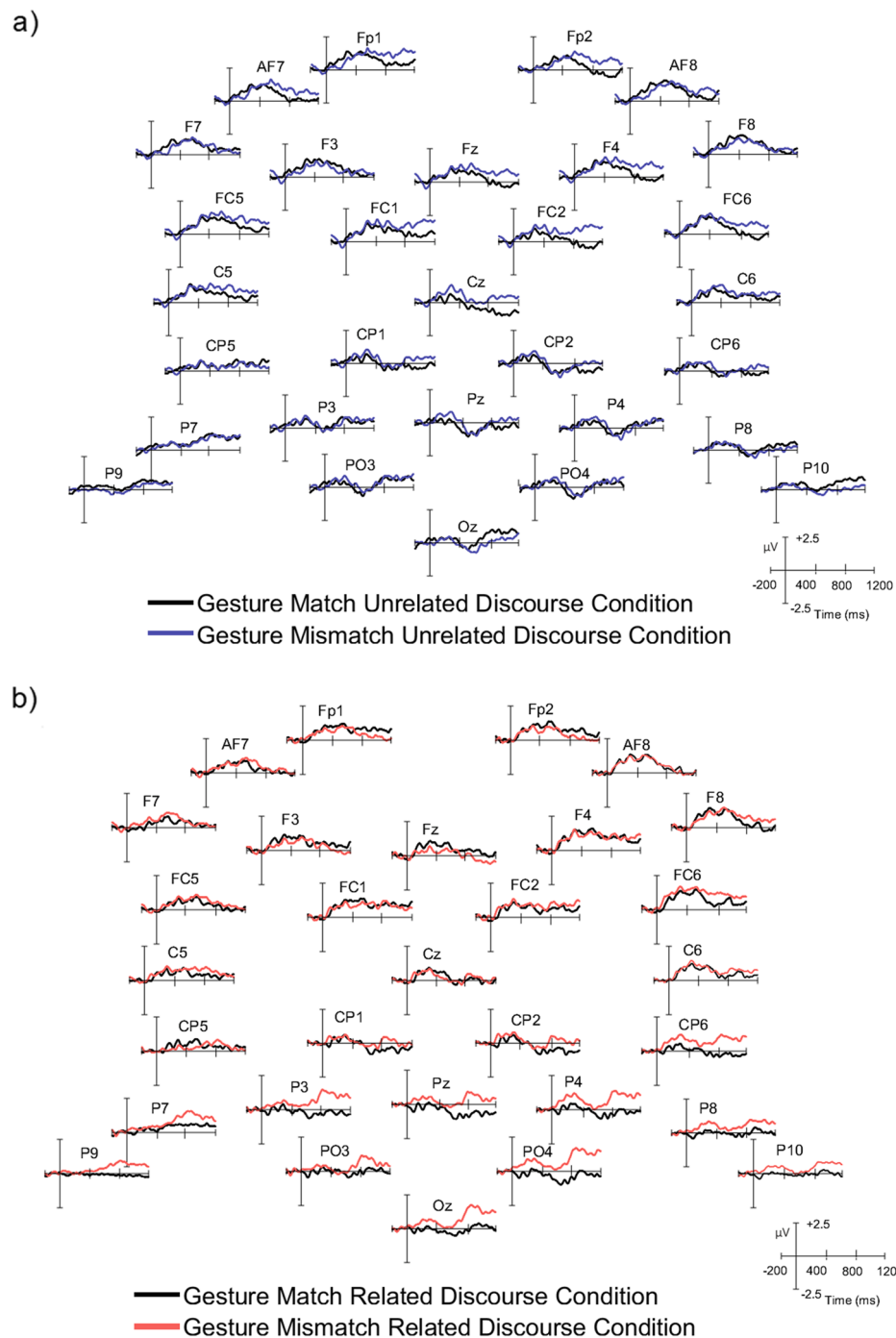
Fig. 8. ERP responses for Gesture Match and Gesture Mismatch at selected electrodes in the Unrelated (panel a) and the Related (panel b) Discourse Conditions, time-locked to the onset of the target verb (the semantic affiliate). Negativity is plotted downwards. Gesture Mismatch in the Related Discourse Condition led to a P600 effect in a 800–900 ms window, while differences in the Unrelated Discourse Condition were not significant.

effects in the typical N400 time-window (300–500 ms). But there were significant effects in the 800–900 ms and the 1000–1100 ms time-windows post verb-onset, as well as trends for the windows surrounding the 800–900 ms window, in particular for the Discourse × Gesture Match × ROI interaction.

In the 800–900 ms window, we found main effects of Gesture Match (*coeff* = 0.96, *SE* = 0.48, *t* = 2.02, *p* = .045) and Discourse × Gesture Match × ROI interactions for the comparisons of the Posterior Right ROI (the reference level) with the Anterior Right ROI (*coeff* = 1.69, *SE* = 0.82, *t* = 2.07, *p* = .039) and the Anterior Left ROI (*coeff* = 1.68, *SE* = 0.82, *t* = 2.05, *p* = .040). Fitting separate Gesture Match × Discourse

models in each ROI showed no significant effects in the Centre or in any of the anterior regions. But we found a significant main effect of Gesture Match (Posterior Left: *coeff* = 1.06, *SE* = .364, *t* = 2.90, *p* = .004; Posterior Right: *coeff* = 0.96, *SE* = 0.38, *t* = 2.51, *p* = .014) and a significant Discourse × Gesture Match interactions in the Posterior Left and Posterior Right regions (Posterior Left: *coeff* = −1.26, *SE* = 0.51, *t* = −2.45, *p* = .013; Posterior Right: *coeff* = −1.27, *SE* = 0.51, *t* = −2.46, *p* = .014). There was also a significant main effect of Discourse in the Posterior Left region (*coeff* = 0.77, *SE* = 0.36, *t* = 2.147, *p* = .032) (see also Table 3 in the Supplementary material). Thus, the three-way interaction was due to the fact that Gesture Match × Discourse interaction was observed only in

Table 2

Results of the mixed effect models for ERPs time-locked to the target verb's (=semantic affiliate's) onset across all time-windows. Posterior Right is reference level for the factor ROI. AR = anterior right ROI, AL = anterior left ROI, PL = posterior left ROI, PR = posterior right ROI; * < .05, + < .1, ns = not significant.

	Time-window				
	200–300	**300–400**	**400–500**	**500–600**	**600–700**
Discourse	ns	ns	ns	ns	ns
Gesture Match	ns	ns	ns	ns	ns
Gesture Match × Discourse	ns	ns	ns	ns	ns
Discourse × AR	ns	ns	ns	ns	ns
Discourse × AL	ns	ns	ns	ns	ns
Discourse × Centre	ns	ns	ns	ns	ns
Discourse × PL	ns	ns	ns	ns	ns
Gesture Match × AR	ns	ns	ns	ns	ns
Gesture Match × AL	ns	ns	ns	ns	ns
Gesture Match × Centre	ns	ns	ns	ns	ns
Gesture Match × PL	ns	ns	ns	ns	ns
Gesture Match × Discourse × AR	ns	ns	ns	ns	ns
Gesture Match × Discourse × AL	ns	ns	ns	ns	ns
Gesture Match × Discourse × Centre	ns	ns	ns	ns	ns
Gesture Match × Discourse × PL	ns	ns	ns	ns	ns
	700–800	**800–900**	**900–1000**	**1000–1100**	**1100–1200**
Discourse	ns	ns	ns	ns	ns
Gesture Match	ns	*	ns	ns	ns
Gesture Match × Discourse	ns	*	ns	ns	ns
Discourse × AR	ns	ns	ns	ns	ns
Discourse × AL	ns	ns	ns	ns	ns
Discourse × Centre	ns	ns	ns	ns	ns
Discourse × PL	ns	ns	ns	ns	ns
Gesture Match × AR	ns	+	+	ns	ns
Gesture Match × AL	ns	+	ns	+	ns
Gesture Match × Centre	ns	ns	+	ns	ns
Gesture Match × PL	ns	ns	ns	ns	ns
Gesture Match × Discourse × AR	+	*	+	ns	ns
Gesture Match × Discourse × AL	ns	*	ns	*	ns
Gesture Match × Discourse × Centre	ns	+	ns	+	ns
Gesture Match × Discourse × PL	ns	ns	ns	ns	ns

the two posterior regions. The waveforms and topographic plot in Fig. 9 indicate that the two-way interaction is due to the fact that the ERP is more positive for Gesture Match than Gesture Mismatch in the Related Discourse condition, but not in the Unrelated Discourse Condition. Thus, we observed a P600 only in the Related Discourse Condition.

In the 1000–1100 ms time-window, we found a significant Discourse × Gesture × ROI interaction for the comparison of the reference ROI (Right Posterior) with the Anterior Right ROI (*coeff* = 1.23, *SE* = 0.86, *t* = 2.02, *p* = .043). Fitting separate Gesture Match × Discourse models in each ROI showed no significant effects (see also Table 4 in the Supplementary Materials). Thus, it is unclear where the three-way interaction stemmed from.

Discussion experiment 2 – ERP experiment

The ERP experiment tested whether discourse information enables an integration of gesture into a recipient's discourse model when the gesture is not synchronised with its semantic affiliate. For the analysis time-locked to the gesture, we had predicted more negative anterior ERPs for the unrelated than related discourse. But, while the ERP waveforms suggested such a difference, our statistical analysis showed no significant difference. Also, as expected, ERPs that were time-locked to the gesture's semantic affiliate, the target verb downstream in the sentence, did not show a gesture mismatch effect in the N400 time-window (between 300 and 500 ms), but showed a P600 effect. More specifically, in the related discourse condition we found a posterior P600 effect (mismatch more positive than match) distributed bilaterally over the posterior sites, in the time-window 800–900 ms post verb onset. This posterior effect was not found in the unrelated discourse condition.

The P600 effect in the 800–900 ms time window for the analysis time-locked to the target verb is unlikely to be a false positive for two reasons. First, the effect was robust both spatially (two anterior ROIs: AR

and AL) and temporally (700–1100 ms) (see Table 2). The crucial Gesture Match × Discourse × ROI interaction was marginally significant in adjacent time windows (700–800, 900–1000 ms, for Gesture Match × Discourse × AR) and significant in a near-by time window (1000–1100 ms, for Gesture Match × Discourse × AL). This indicates that it was not a random "blip" in the data that appeared for a short period of time at some arbitrary electrodes. Second, the scalp distribution of our P600 was consistent with many previous studies (e.g., Van Petten & Luka, 2012; Brouwer, Crocker, Venhuizen, & Hoeks, 2017; Brouwer et al., 2012; Brouwer & Hoeks, 2013).

Lack of anterior negativity for gesture-locked ERPs

We did not find any significant effect of preceding discourse on gesture interpretation in the gesture-locked ERP analysis. This, however, does not mean that participants ignored the preceding discourse because the two discourse conditions made a difference in the verb-locked ERPs. The lack of gesture-locked ERPs may be because the gesture is not synchronised with a semantic affiliate and the preceding discourse only loosely constrained gesture interpretation. Thus, the semantic relationship between preceding discourse (e.g., "Some of the strawberries in the garden were already ripe.") and a gesture (e.g., a gesture enacting picking) may become apparent at different time points across items and across participants. That is, the integration process may not be tightly time-locked to the gesture onset, and thus may not appear as a clear ERP.

P600 effect in verb-locked ERPs

Our results for the P600 in the related discourse condition suggests that the related discourse enabled the participants to construct a relatively concrete meaning for a gesture. When the semantic affiliate (the verb) downstream in the sentence was integrated into the discourse model and the gesture's interpretation did not match the semantic
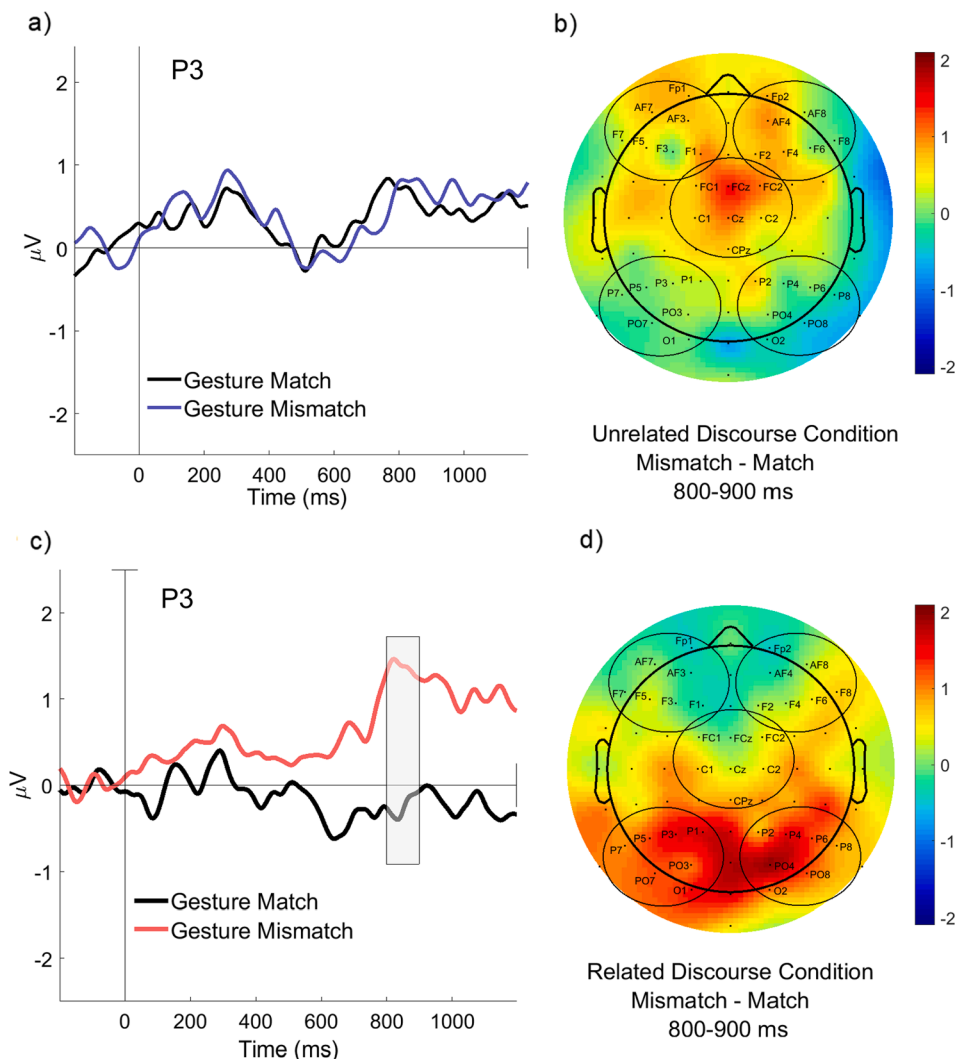
Fig. 9. ERPs time-locked to the verb for the sample electrode P3 in both the Unrelated (panel a) and Related (panel c) Discourse conditions, showing the difference between Gesture Match and Gesture Mismatch. The grey shaded area indicates the significant time-window (800–900 ms, time-locked to the onset of the target-verb, i. e. the semantic affiliate). Topographical maps depict ERP differences between matching and mismatching gestures (Mismatch minus Match) in the Unrelated Discourse Condition (panel b) and the Related Discourse Condition (panel d) for the 800–900 ms time-window. Black dots indicate electrode sites and circles the ROIs included in the statistical analysis.

affiliate, the gesture's meaning was reanalysed in the attempt to re-incorporate it into the discourse model. This process led to a P600 effect.

The absence of a P600 effect in the unrelated discourse condition suggests that the verb was processed very similarly after a matching or mismatching gesture. This may be because the interpretation of the gesture was not concrete enough to trigger a reanalysis process. Alternatively, the interpretation of the gesture may have been reasonably concrete but too diverse, and thus even in the match condition, the interpretation of the gesture often diverged from the target verb meaning and made the contrast between the match and mismatch condition small.

General discussion

The current study investigated two questions: 1) whether preceding verbal discourse can constrain how recipients interpret an iconic gesture that does not synchronise with (i.e., that starts and ends before) its semantic affiliate in speech, and 2) whether this discourse-based constraint on gesture interpretation can influence how the recipients integrate the iconic gesture's semantic affiliate, which is located later in the sentence, into the discourse model.

As for the first research question, we investigated how recipients interpreted an iconic gesture (e.g., a gesture that enacts picking) in situations where the verbal discourse preceding the iconic gesture was semantically related (e.g., "Some of the strawberries in the garden were

already ripe"; the Related Discourse Condition) versus unrelated (e.g., "At the beginning of the week the weather was dreadful"; the Unrelated Discourse Condition). In the behavioural experiment (Experiment 1), when preceding verbal discourse was semantically related to the gesture, the recipients found it easier to interpret the gesture and they tended to agree more with each other on the gesture interpretation (i.e., less diversity of interpretations across recipients). However, the interpretations were still wide-ranging across the recipients, suggesting that the gesture interpretations were not fully disambiguated. The ERP experiment (Experiment 2), however, did not find any significant gesture-locked ERP difference between the Related and Unrelated discourse conditions. These results indicate that preceding discourse can constrain the interpretation of an iconic gesture that is not synchronised with its semantic affiliate. However, this process may not be tightly time-locked to the gesture (e.g., varying from item to item) and/or this process may not constrain interpretations sufficiently to create enough difference between the Related and Unrelated Discourse conditions, which does not lead to strong ERP effects for the discourse manipulation.

As for the second research question, we investigated how recipients judged the semantic congruency between an iconic gesture (e.g., a gesture that enacts picking) and a target word, which was downstream in the sentence. The target word was either the gesture's semantic affiliate (e.g., "to pick"; the Gesture Match Condition) or not (e.g., "to water", the Gesture Mismatch Condition). In the behavioural experiment (Experiment 1), the recipients judged their interpretation of the

iconic gesture (e.g., a gesture that enacts picking) to match more strongly with the semantic affiliate (e.g., "picking") when the verbal discourse preceding the gesture was semantically related to the gesture (the Related Discourse Condition) than when it was not (the Unrelated Discourse Condition) (Question 3). Unsurprisingly, the recipients judged their gesture interpretations to match better with semantic affiliates (e. g., "to pick") than non-semantic affiliates (e.g., "to water"). More importantly, for non-semantic affiliates, the congruency judgement did not differ between the two discourse conditions. That is, the semantically related preceding discourse constrained the interpretation of gesture in the right direction towards the meaning of the semantic affiliate. The ERP analysis time-locked to the verb (i.e. the semantic affiliate, Experiment 2) found a converging result. A more positive P600 component was found for non-semantic affiliates (the Gesture Mismatch Condition) than semantic affiliates (the Gesture Match Condition) when the preceding discourse was semantically related to the gesture (the Related Discourse Condition), but not when the preceding discourse was semantically unrelated (the Unrelated Discourse Condition). This indicates that when preceding discourse constrained the interpretations of a gesture and made it more concrete (but did not completely disambiguate it), the recipient re-analysed gesture interpretation upon encountering a mismatching target word.

Taken together, we conclude that an iconic gesture *can* be integrated into a discourse model even when the gesture precedes and terminates before its semantic affiliate, but only when discourse preceding the gesture is semantically related to the gesture. In this case, the discourse can constrain the interpretation of the gesture to some extent. We further conclude that recipients *dynamically update* how a gesture is integrated into a discourse model as the discourse unfolds. Gesture interpretation that has not been fully specified (disambiguated) can be re-analysed as more relevant information enters the discourse. This conclusion provides a novel understanding of how speech and gesture are integrated in situations that are often observed in natural settings, where gestures are embedded in verbal discourse, and of how a gesture's contribution to a discourse model becomes gradually clearer as the discourse unfolds (e.g., Fritz, 2018; Kita, 2000, Kita, Alibali, & Chu, 2017).

The conclusion from the current study and insights from previous research lead to a more general theoretical framework for how recipients integrate speech and gesture. As recipients process incoming discourse, they build a discourse model representing the interpretation of the discourse. Recipients use information from all sources as soon as they become available, including preceding verbal discourse (e.g., van Berkum, 2009), co-speech gesture (e.g., Özyürek et al., 2007) plus presumably other nonverbal communicative behaviours (such as facial expressions), real-world knowledge (Hagoort et al., 2004), and common ground with a communication partner (Clark, 1996). Interpretations of gesture and speech mutually constrain each other (Sekine, Sowden, & Kita, 2015) in a dynamic way.

When a recipient sees a gesture in a discourse setting, one of the following happens, depending on how specific their interpretation of the gesture is, based on various contextual information, including preceding verbal discourse. 1) The first scenario is when the recipient can derive a specific interpretation of a gesture immediately (e.g., because the gesture is synchronised with its semantic affiliate in speech or the preceding discourse provides a highly specific context). If the gesture interpretation is compatible with the discourse model, the interpretation adds definitive information to the discourse model and it can facilitate/ prime decoding of concurrent and following speech (e.g. lexical retrieval of the semantic affiliate) (e.g., Wu & Coulson, 2010). If the gesture interpretation is *not* compatible with the discourse model, N400 effects can be observed at the gesture (Wu & Coulson, 2005; Özyürek et al., 2007). The speech following the gesture may include information incompatible with the gesture and the discourse model; in that case an N400 can be observed (Holle & Gunter, 2007). 2) The second scenario is when the recipients can only derive an ambiguous interpretation of a

gesture (compatible with multiple distinct interpretations). In this case, the ambiguous interpretation enters the discourse model, and it can loosely constrain the interpretation of concurrent or subsequent speech. As the discourse unfolds, the gesture interpretation can become more specific and gets more concretely incorporated into the discourse model. However, if the subsequent speech provides information incompatible with the vague gesture interpretation, then the gesture interpretation is re-analysed and gets re-incorporated into the discourse model (this process is reflected in a P600, as seen in the current study; cf. Brouwer et al., 2012). 3) The third scenario is when the recipient can only make very little sense of a gesture. In this case, the information about the gesture does not enter the discourse model. It is possible that some uninterpreted visual information about the gesture may remain in memory (outside of the discourse model) for a short period of time, and this could be interpreted with information provided subsequently as long as the visual representation is still available. The level of ambiguity of gestures differs across these three scenarios: it is least ambiguous in (1) and most ambiguous in (3). In terms of ERPs, more ambiguous gestures might potentially elicit an anterior negativity when encountered (e.g., Hagoort & Brown, 1994; Federmeier, Segal, Lombrozo, Kutas, 2000; van Berkum, 2009; Dyck & Brodeut, 2015). In the current study, we assume that this process for ambiguous communication signals was not properly time-locked to the gesture onset; thus, we did not find the significant anterior negativity.

The current results extend the previous literature on integration of speech and iconic gesture in four significant ways. First, in terms of gesture-speech synchrony, the current study indicates that gesture processing is more flexible than is often assumed (cf. McNeill, 1992). Habets et al. (2011) concluded that speech and gesture cannot be integrated with each other when an iconic gesture precedes (and does not overlap) with its semantic affiliate in speech. However, Habets and colleagues' study used highly de-contextualised stimuli, consisting of pairs of a single iconic gesture and a single word. Our results indicate that when discourse supports gesture interpretation, gestures that do not synchronise with their semantic affiliate can be integrated with verbal discourse.

Second, in terms of semantic gesture-speech integration, the current study shows that gestures can be integrated with speech even when the gesture interpretation is not fully disambiguated by speech at the point of gesture production, and gets disambiguated much later in the sentence. This extends previous research on semantic integration of speech and gesture, which mostly investigates the situation where gesture co-occurs with (or immediately followed by) words that would fully disambiguate the gesture interpretation (e.g., Kelly, Kravitz, & Hopkins, 2004; Kelly, Özyürek, & Maris, 2010; Wu & Coulson, 2010; Özyürek, et al., 2007; Holle, & Gunter, 2007; Sekine et al., 2015). One notable exception in the literature are the studies by Obermeier and Gunter (2015) and Obermeier et al. (2011), which we will discuss in more depth below.

Third, the current study also shows that recipients can loosely constrain gesture interpretation using the preceding discourse without fully disambiguating it. This extends the finding in the literature that recipients can use the preceding discourse to fully disambiguate gesture interpretation (Özyürek et al., 2007; Sekine & Kita, 2015, 2017; Gunter & Weinbrenner, 2017).

Fourth, the current study shows that iconic gesture interpretation is dynamically updated as the discourse unfolds. This was shown by the P600 effect in Experiment 2, which suggested that gesture interpretation is re-analysed when the target word later in the sentence was incompatible with the initial gesture interpretation. So far, only one gesture study has reported a similar P600 (Gunter & Weinbrenner, 2017). In their study, participants watched an actor placing two referents (Shakespeare vs. Goethe) left vs. right in gesture space with an abstract pointing gesture. Later in the discourse, an abstract pointing gesture (i.e., pointing to the left or right) was used synchronous with Shakespeare or Goethe in speech, and indicated a location that matched or mismatched the previously

established locations for the referent (e.g., Goethe was on the right, Shakespeare on the left). A P600 effect was elicited when the gesture did not match the target word. The authors argued that participants tried to reanalyse meanings as gesture space when the target word was integrated into a discourse model, which elicited a P600 gesture mismatch effect. Thus, the current results dovetail with Gunter and Weinbrenner's finding, and extend it to iconic gestures whose interpretation becomes more specific as the discourse unfolds.

Together with previous findings, the current study sheds light on what types of gesture-speech semantic incongruency elicits an N400 as opposed to a P600. Following Gunter and Weinbrenner (2017) and Gunter et al.'s (2015) account of a P600 elicited by abstract pointing gestures, we expand on Brouwer and colleagues' (Brouwer et al., 2012) account of the N400 and P600. Namely, we argue that the N400 reflects context driven meaning construction (for gesture) and/or semantic lexical retrieval, whilst the P600 indicates post-semantic integration into a discourse model. When gesture and semantic affiliate co-occur and the congruent preceding discourse and the semantic affiliate fully disambiguate gesture interpretation (e.g., Holle & Gunter, 2007; Özyürek et al., 2007), the incongruent semantic affiliate should elicit an N400 effect, reflecting difficulty in meaning construction of the gesture or semantic lexical retrieval. In contrast, when the gesture does not co-occur with its semantic affiliate, but the related preceding discourse (loosely) constraints gesture interpretation, the information later in the discourse (e.g., the gesture's semantic affiliate downstream the sentence) that is incompatible with the gesture interpretation should elicit a P600, reflecting re-analysis of the gesture interpretation.

Furthermore, our results qualify Obermeier and Gunter's (2015) proposal on speech-gesture integration processes when gestures do not synchronise with (i.e., precede) their semantic affiliates. They proposed (based on findings in Obermeier et al., 2011, and Obermeier & Gunter, 2015) that when such a gesture co-occurs with a content word, an N400 effect should be observed at the (semantically matching or mismatching) target word downstream the sentence. But, when the gesture co-occurs with a function word, such an N400 effect should not be observed. In the current study, though the gesture synchronised with a content word, we did not find an N400 gesture mismatch effect on the semantic affiliate but we found a P600 effect instead. We argue that in both studies by Obermeier and colleagues (Obermeier & Gunter, 2015; Obermeier et al., 2011), an N400 effect was elicited at the downstream target word because the gesture fully disambiguated the co-occurring content word and/or preceding discourse, and the fully specified gesture interpretation primed the semantic affiliate. Based on the current findings, we would expect a P600 effect when a gesture synchronises with a function word and is then disambiguated by the target word, as in Obermeier et al.'s studies, due to a reanalysis of the gesture interpretation. However, neither of the two studies investigated the time window appropriate for the P600.

The current study used a particular pattern of synchronisation between iconic gesture and speech in the stimuli, and it is an important topic for future studies to investigate how representative the synchronisation pattern is in natural discourse. In general, there is very limited quantitative information about how people synchronise speech and iconic gestures (notable exceptions include Fritz, 2018; Morrel-Samuels & Krauss, 1992; Nobe, 2000); thus, at this point, it is difficult to assess what patterns of synchronisation is representative.

The present study had a single word as a semantic affiliate of each gesture. This raises the question as to whether the processing would be substantially different if we had longer multi-word stretches as a semantic affiliate. There has been ERP evidence for semantic integration with multiple words (e.g., integration with metaphorical sentences, Cornejo et al., 2009; Ibáñez et al., 2011). We speculate that the processing will not qualitatively change for multi-word semantic affiliates: the semantic affiliate that is not compatible with a loosely constrained gesture interpretation would trigger a re-analysis of the gesture interpretation. One difference may be that the processing would be less well

time-locked to a particular word (e.g., the first word) in the semantic affiliate, which means the P600 effect may be less clear. In general, the picture that emerges from the present study and the literature is that recipients make sense of speech and gesture as much as possible with the information available at a given moment in speech, gesture and context.

The two experiments in the current study differed in that only the behavioural experiment (Experiment 1) explicitly asked participants to interpret the gestures. This might have had an effect on the results. It might be that explicit gesture interpretations in the EEG experiment have been more tightly time-locked to the gesture if its task had forced participants to semantically interpret the gestures. In this case, we might have found the predicted anterior negativity effect of our discourse manipulation on ERPs time-locked to the gesture. We might also have found larger P600 effects time-locked to the semantic affiliate (the verb). However, we found evidence that the participants in the ERP experiment interpreted the gestures in relation to the preceding discourse: discourse made a difference in the processing of the gestures' semantic affiliates (P600 on the verbs). Also, previous studies suggest that recipients cannot help but to interpret gestures when presented with gesture-speech stimuli even if the meaning of the gesture is not relevant to the task (Kelly, Özyürek, & Maris, 2010).

To conclude, the current study sheds light on how recipients integrate information from speech and iconic gesture in a realistic situation where the meaning of a gesture becomes gradually clear as discourse unfolds. Recipients use preceding discourse to constrain gesture interpretation even if the interpretation is vague, and then dynamically update the interpretation when more disambiguating information comes in later in the discourse. We hope that this study will lay a foundation for future studies on the subtle and dynamic semantic interplay between gesture and speech in discourse.

CRediT authorship contribution statement

Isabella Fritz: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing - original draft, Writing - review & editing. **Sotaro Kita:** Conceptualization, Methodology, Supervision, Writing - original draft, Writing - review & editing. **Jeannette Littlemore:** Conceptualization, Supervision, Writing - review & editing. **Andrea Krott:** Conceptualization, Supervision, Methodology, Writing - review & editing.

Acknowledgment

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2020.104191.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fittinglinearmixed-effectsmodels usinglme4.2015,67(1),48.doi:https://doi.org/10.18637/jss.v067.i01.

Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 18*(4), 438–462. https://doi.org/10.1177/0261927X99018004005.

Beattie, G., & Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology, 93*(Pt 2), 179–192.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A Neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science, 41*, 1318–1352. https://doi.org/10.1111/cogs.12461.

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.

Brouwer, H., & Hoeks, J. (2013). A time and place for language comprehension: Mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience, 7*(758). https://doi.org/10.3389/fnhum.2013.00758.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., López, V., & Núñez, R. E. (2009). Gesture and metaphor comprehension: Electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain and Cognition, 70*(1), 42–52. https://doi.org/10.1016/j.bandc.2008.12.005.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture* (pp. 248–311). Cambridge: Cambridge University Press.

Dyck, M., & Brodeur, M. B. (2015). ERP evidence for the influence of scene context on the recognition of ambiguous and unambiguous objects. *Neuropsychologia, 72*, 43–51. https://doi.org/10.1016/j.neuropsychologia.2015.04.023.

Federmeier, K. D., Segal, J. B., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. *Brain, 123*, 2552–2566.

Feyereisen, P., Vandewiele, M., & Dubois, F. (1988). The meaning of gestures – what can be understood without speech. *Cahiers de Psychologie Cognitive, 8*(1), 3–25.

Fritz, I. (2018). *How Gesture and Speech Interact during Production and Comprehension* (PhD Thesis). University of Birmingham.

Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help Us Think*. Cambridge, MA: Belknap of Harvard UP.

Gunter, T. C., & Weinbrenner, J. E. D. (2017). When to take a gesture seriously: On how we use and prioritize communicative cues. *Journal of Cognitive Neuroscience, 1–12*. https://doi.org/10.1162/jocn_a_01125.

Gunter, T. C., Weinbrenner, J. E. D., & Holle, H. (2015). Inconsistent use of gesture space during abstract pointing impairs language comprehension. *Frontiers in Psychology, 6* (80). https://doi.org/10.3389/fpsyg.2015.00080.

Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neurosciences, 23*, 1845–1854.

Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture: Implications for gesture classification and function. *Journal of Language and Social Psychology, 23* (2), 204–214. https://doi.org/10.1177/0261927x04263825.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*(6), 883–899. https://doi.org/10.1162/089892903322370807.

Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Erbaum: Hillsdale, NJ.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438–441. Retrieved from <Go to ISI>://000220845400048.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience, 19*, 1175–1192.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137* (2), 151–171.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118–135.

Ibáñez, A., Manes, F., Escobar, J., Trujillo, N., Andreucci, P., & Hurtado, E. (2010). Gesture influences the processing of figurative language in non-native speakers: ERP evidence. *Neuroscience Letters, 471*(1), 48–52. https://doi.org/10.1016/j.neulet.2010.01.009.

Ibáñez, A., Toro, P., Cornejo, C., Urquina, H., Manes, F., Weisbrod, M., & Schroder, J. (2011). High contextual sensitivity of metaphorical expressions and gesture blending: A video event-related potential design. *Psychiatry Research, 191*(1), 68–75. https://doi.org/10.1016/j.pscychresns.2010.08.008.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 253–260. https://doi.org/10.1016/S0093-934X(03)00335-3.

Kelly, S. D., Manning, S. M., & Rodak, S. (2008). Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education. *Language and Linguistics Compass, 2*(4), 569–588. https://doi.org/10.1111/j.1749-818X.2008.00067.x.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*(2), 260–267. https://doi.org/10.1177/0956797609357327.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162–185). Cambridge: Cambridge University Press.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review, 124* (3), 245–266. https://doi.org/10.1037/rev0000059.

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology, 61*(5), 743–754. https://doi.org/10.1037/0022-3514.61.5.743.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.

Kutas, M., Federmeier, K. D., Staab, J., & Kluender, R. (2007). Language. In G. Berntson, J. T. Cacioppo, & L. G. Tassinary (Eds.), *Handbook of Psychophysiology* (3 ed., pp. 555–580). Cambridge: Cambridge University Press.

Lee, C.-L., & Federmeier, K. D. (2009). Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. *Journal Of Memory And Language, 61*(4), 538–555. https://doi.org/10.1016/j.jml.2009.08.003.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92*, 350–371.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago: The University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago: The University of Chicago Press.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 615–622. https://doi.org/10.1037/0278-7393.18.3.615.

Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech? In D. McNeill (Ed.), *Language and Gesture* (pp. 186–198). Cambridge: Cambridge University Press.

Nunez, M. D., Nunez, P. L., & Srinivasan, R. (2016). Electroencephalography (EEG): Neurophysics, experimental methods, and signal processing. *Handbook of Statistical Methods for Brain Signals and Images*, 175–197.

Obermeier, C., & Gunter, T. C. (2015). Multisensory integration: The case of a time window of gesture-speech integration. *Journal of Cognitive Neuroscience, 27*(2), 292–307. https://doi.org/10.1162/jocn_a_00688.

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.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59* (3), 494–522. https://doi.org/10.1006/brln.1997.1793.

Otten, M., & van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: prediction or priming? *Discourse Processes, 45*(6), 464–496. https://doi.org/10.1080/01638530802356463.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org.

Schegloff, E. A. (1984). On some gesture's relation to talk. In M. A. J. Heritage (Ed.), *In Structures of Social Action: Studies in Conversation Analysis* (pp. 266–296). Cambridge: Cambridge University Press.

Sekine, K., & Kita, S. (2015). Development of multimodal discourse comprehension: Cohesive use of space by gestures. *Language, Cognition and Neuroscience, 30*(10), 1245–1258. https://doi.org/10.1080/23273798.2015.1053814.

Sekine, K., & Kita, S. (2017). The listener automatically uses spatial story representations from the speaker's cohesive gestures when processing subsequent sentences without gestures. *Acta Psychologica, 179*, 89–95. https://doi.org/10.1016/j.actpsy.2017.07.009.

Sekine, K., Sowden, H., & Kita, S. (2015). The development of the ability to semantically integrate information in speech and iconic gesture in comprehension. *Cognitive Science, 39*(8), 1855–1880. https://doi.org/10.1111/cogs.12221.

van Berkum, J. J. A. (2008). Understanding sentences in context. *Current Directions in Psychological Science, 17*(6), 376–380. https://doi.org/10.1111/j.1467-8721.2008.00609.x.

van Berkum, J. J. A. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Palgrave Macmillan.

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.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015.

Willems, R. M., van der Haegen, L., Fisher, S. E., & Francks, C. (2014). On the other hand: Including left-handers in cognitive neuroscience and neurogenetics. *Nature ReviewsNeurosci, 15*(3), 193–201. https://doi.org/10.1038/nrn3679.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42*(6), 654–667. 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 & Review, 14*, 57–63.

Wu, Y. C., & Coulson, S. (2010). Gestures modulate speech processing early in utterances. *NeuroReport, 21*, 522–526. https://doi.org/10.1097/WNR.0b013e32833904bb.

Özyürek, A. (2001). What do speech-gesture mismatches reveal about language specific processing? A comparison of Turkish and English. Proceedings of the 27th Annual Meeting of the Berkeley Linguistics Society (BLS27), 449-456.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society 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.