The effects of discourse predictability and iconic gestures on sentence comprehension

Adam Psomakas

s1063846

August 2021

LET-TWM400: Master's Thesis

MA Linguistics: General Linguistics

Supervision: Dr. Florian Hintz (primary assessor)

Second reader: Dr. Falk Huettig

Radboud Universiteit

Table of contents

Abstract

An increasing number of studies has investigated language comprehension in natural settings. In these types of settings, various visual cues, such as iconic gestures, are conveyed alongside the auditorily transmitted speech, which serve as contributing sources of information. Iconic gestures have been shown to provide the addressee with valuable information about actions and objects that are referred to in the speech. Importantly, these iconic gestures are often semantically coupled and temporally aligned with spoken entities they denote, thus supplementing speech, and enhancing comprehension from the addressee's end. Previous literature has shown that comprehenders also frequently engage in predictive processing by exploiting cues from the given discourse context. To date, no study has yet investigated the interaction of iconic gestures and discourse predictability when studying their contribution to sentence comprehension. The present study aimed to fill that gap. EEG was recorded while participants were presented with sentences, where a sentence-initial context was either predictable or non-predictable, and an iconic or a control gesture was coupled with the sentence-final target word. The study focused primarily on the N400 component, to examine main effects of discourse predictability and gesture iconicity, as well as the effect of the interplay of these two factors, with respect to the ease of processing of the target word. Independent main effects of gesture iconicity and discourse predictability were indeed obtained, showing that sentence comprehension can be facilitated by a predictable discourse and an iconic gesture individually. Moreover, an exploratory analysis on a later time window showed that these main effects were sustained, indicating conserved efficiency with respect to the target word processing. Finally, no significant effect was found when the two factors interacted. This later finding poses some interesting questions about the psycholinguistic processing within communicative situ (Holler & Levinson, 2019), which are thoroughly addressed and discussed.

Keywords: multimodal communication, comprehension, discourse predictability, gesture iconicity, semantic integration

Introduction

Many studies in cognitive science have investigated the notion of predictive processing in the human mind, and there is now compelling evidence that prediction plays an important role in cognitive processing. Studies on prediction have helped substantially in shaping new cognitive models that can illustrate more accurately how the mind functions. In the words of Clark (2013), the human brain, ultimately, might constitute a prediction machine. Predictive processing can be thought of as a way of structuring considerable and reliable expectations, with regard to potential continuations that will unfold in the near future, based on the given context, as well as on memory (Bar, 2007). For the auditory domain, Bendixen et al. (2009) conducted an EEG study, focusing on differential brain responses to predictable versus non-predictable tones. Interestingly, as inferred by the elicited electrical activity, the findings supported the authors' claim that the auditory system engages in predictive processing.

In the same direction, Bar (2009) highlighted the importance of predictive processing when it comes to visual perception as well. As he stated, when coming across frequently occurring objects, perceived through vision (and other senses), we tend to recognise and extract certain patterns and regularities. In turn, these regularities are used to link, through analogy, any new input to the already existing representations that reside in our memory. Finally, this newly acquired associated information allows us to make reliable predictions that can enhance our future interactions. This is also evident when focusing on language comprehension. Van Berkum et al. (2005) conducted an EEG study, pointing out how crucial a given sentence discourse can be, as it can provide valuable information that, in turn, can be utilized to make well-grounded predictions about the upcoming input.

As described, language processing is tightly linked to the visual and the auditory perception. Moreover, the topic of language processing and comprehension, and how predictive processing is involved, either in the auditory or the visual modality, has been widely explored through many studies, like the one of Van Berkum (2005), which will be discussed below. However, little is known about how language processing operates in an audio-visual (i.e., multimodal) context. Thus, it is the present study's goal to provide insights about the potentially dynamic and multifaceted nature of sentence comprehension, where a mechanism, such as predictive processing, plays a pivotal role.

Recent frameworks have highlighted the fact that language communication, particularly in natural face-to-face conditions, entails various articulators, both vocal and non-vocal (e.g., hands), in the auditory and the visual modality (Holler & Levinson, 2019). Even though people primarily speak to accomplish their communicative goals, they consistently accompany their speech with intended eye-gazes, head movements, and, especially, hand gestures. In fact, these additional lines of communication appear to enhance language processing from the addressee's perspective to a substantial extent. Messages that are conveyed in a multimodal context are processed with significantly greater ease, compared to messages that are confined to a unimodal setting. One of the most widely used type of gestures is the that of iconic gestures, often coupled with a part of speech, as they both provide semantic information that match with each other (McNeill, 1985; McNeill, 1992). Iconic gestures are remarkably capable of transmitting concrete information about actions and objects.

An intriguing question that has not been addressed by previous research is how, and to what extent, discourse and gestures may jointly contribute to sentence comprehension. Thus, the present study aimed to investigate whether and how predictive processing interacts with processing iconic gestures. In other words, the study seeks to discover if, and to what extent, these two factors, discourse predictability and gesture iconicity, can lead to the formation of a unified mental representation, with the aim of rendering the processing of the upcoming information more efficient and effortless, thus facilitating sentence comprehension.

Prediction during language comprehension

As indicated above, prediction constitutes a mechanism that is integral to human thought and deeply associated with perception. Thus, it is only natural to suggest that predictive processing is strongly involved with a mental process as central and multi-layered as language. There is a great number of studies that have focused on investigating the mechanisms underlying predictive processing during language comprehension. Many of them have conducted their experiments using electroencephalography (EEG), by focusing on the so-called event-related potentials (ERPs). ERPs are measured brain responses, which constitute the outcome of specific motor, cognitive or sensory events, and can be thought of as event-related voltage changes, which take place during the continuing EEG activity (Luck, 2014). Importantly, they are time-locked to the events that have triggered these changes (Kropotov, 2016). Studies on predictive language comprehension have predominantly focused on the N400 component, a centro-parietal negativity that occurs approximately 400 ms after stimulus onset (Kutas & Federmeier, 2011). The reason why this specific ERP has been the focus of most of these studies, is because it is considered to be a fairly and substantially accurate index of semantic and lexical processing during language comprehension.

One of the studies, where EEG was used to investigate predictive processing, is the one by Federmeier and Kutas (1999). In this study, electrical activity was recorded, while participants were exposed to pairs of sentences. The first sentence provided a context that varied in terms of contextual strength (high versus low contextual constrain), whereas the second sentence ended with a target word, a noun. Critically, the target word varied semantically in terms of cloze probability with respect to the context-induced expectation, meaning that it would be either an expected exemplar, an unexpected one from the same semantic category, or an unexpected one from a different semantic category. Expected exemplars elicited a smaller N400 compared to unexpected ones, indicative of how predictive processing is tightly involved with language processing and can enhance comprehension.

Apart from focusing on the semantic level, prediction in language comprehension has been investigated at other linguistic levels as well. For instance, Wicha et al. (2004) investigated to what extent addressees can utilize cues of gender agreement to predict the upcoming information. By focusing on the N400, as well as on the P600 component (another ERP that reflects the processing of semantic and morphosyntactic violations), time-locked to an article and noun at the end of a sentence, the authors found that gender disagreement and semantic incongruency elicited both a more negative N400, as well as a more positive P600. These findings further demonstrate that semantic and syntactic cues contribute to forming predictions about future information, as well as to building sentence meaning based on the given discourse context.

A similar finding is that of the study by Van Berkum, Brown, Zwitserlood, Kooijman and Hagoort (2005). EEG was recorded, while Dutch-speaking participants were presented with sentences containing an initial context that was predictable towards a certain sentence-final noun. The realized noun would be either the prediction-consistent one or a prediction-inconsistent one. Critically, these two nouns differed in syntactic gender, and the authors added a preceding gender-marked adjective, which was grammatically congruent with the realized noun. Interestingly, the — gender-wise — prediction-inconsistent adjective elicited a larger N400 amplitude compared to the prediction-consistent one, a difference that was not observed when, in a second experiment, the sentence-initial context was non-predictable.

These findings, not only support the notion of predictive processing in language comprehension, but also suggest that parsing operations are sensitive to prediction.

Earlier research has also investigated the activation of phonology during predictive processing. In an EEG study conducted by DeLong et al. (2005), ERPs were recorded while participants were exposed to sentences, such as *'The day was breezy so the boy went outside to fly...'*, which would end with either a highly predictable phrase, namely '*a kite*' or with a less predictable phrase, like *'an airplane'*. The revealing finding was that the different N400 effect was found at the words '*a'* or '*an'* (before the presentation of the target word), with a reduced N400 in the case of *'a'*, since the preceding context made participants predict the word '*kite'* (starting with a consonant) to follow, rather than '*airplane'* (starting with a vowel). The authors interpreted this finding as clear evidence of predictive processing at the level of phonology.

Another study that is line with this claim is that of Bendixen, Schröger and Winkler (2009). In an EEG study, participants were exposed to sound sequences containing an omitted tone, whose predictability was manipulated. Interestingly, event-related electrical brain activity differed, depending on whether the tone that was omitted was predictable or not. This finding demonstrated how the auditory system, by extracting regularities from the input, pre-activates certain neural circuits, thus leading to sequential predictions about the upcoming input.

To summarize, the literature appears to be congruent with respect to the claim that predictive processing constitutes a fundamental and integral part of sentence processing and language comprehension. As Pickering and Garrod (2013) have argued, prediction appears to play an essential role in language comprehension. An intriguing question that arises is whether, and to what extent, the ease of sentence comprehension is influenced by non-verbal signals during spoken discourse. The question becomes even more interesting when putting it in a multimodal perspective, where, apart from the auditory channel, the interlocutors utilize the visual channel as well, to enhance the communicative process even more. A good example of such communicative visual signals are iconic gestures, a particularly meaningful type of manual gestures, able to convey important semantic information along the spoken discourse.

The contribution of iconic gestures to language comprehension

When it comes to face-to-face interaction, interlocutors are engaged in a communicative process that is multimodal, as they are attentive to both the auditory and the visual channel. Moreover, meaningful messages are conveyed and perceived through both channels. This means that, in these natural conditions, the addressee is anything but surprised when encountering a spoken word that is accompanied by a meaningful eye-gaze, a ‘facial shrug’ or a manual gesture. Manual gestures constitute a prominent type of visually conveyed signals, since they can encode a variety of different messages, due to the hands and arms’ unique physiology, composition, and kinematic potential. One of the most widely used manual gestures are iconic gestures, which generally convey information about specific actions and objects, which the spoken discourse is usually also referring to.

One of the first authors that contributed towards this shift and new perspective, was McNeil (1992), who, in his review, examined the findings of prior research spanning over a period of 10 years, on the topic of gestures. As he argued, gestures should not be thought of as mere hand and arm movements, but rather as “symbols that exhibit meanings”. In his view, “gestures and speech are integral parts of a single process”. However, it was not until recently that it became possible for researchers to investigate the multimodal nature of language in a more methodological way. More specifically, affordable computers, the ability to utilize video recording, and more advanced editing techniques allowed for such reliable experiments to be conducted.

Since then, a growing body of evidence has been further highlighting the significance of gestures contributing to language comprehension. As mentioned before, among the different types of manual gestures, iconic gestures stand out due to their ability to carry remarkably concrete semantic information which echoes the meaning of the co-occurring speech. That is, iconic gestures are usually coupled with spoken entities, in terms of both temporal overlap and semantic association, as they provide — either supplementary or redundant — semantic information about the concept referred by its spoken semantic affiliate. This coupling of speech and co-speech gestures has often been argued to serve as a means of enhancing the process of comprehension (Wu & Coulson, 2007).

This argument is directly associated with the claim that a multimodal condition of human interaction is substantially more effective than a unimodal one. A cluster of information, deriving from meaningful signals that are transmitted through both the auditory and visual channel, is far more potent than a unit of information that results exclusively from a visual or an auditory signal. As Holler and Levinson (2019) argue, this cluster of information, a gestalt, constitutes a holistic percept, capable of rendering the integration of the contained information remarkably more efficacious and effortless, compared to a signal deriving from a single articulator.

Many EEG studies have been conducted in the past years to further strengthen this argument, most of which have done so by implementing mismatch paradigms. An example of such a study is the one carried out by Kelly, Kravitz and Hopkins (2004), who were interested in investigating the potential of hand gestures influencing ERPs to speech, as well as the time course of this influence. EEG was recorded while participants were exposed to audio-visual segments of a word accompanied by a gesture. In the videos, the person, sitting behind two objects (a glass and a dish), would utter a word that corresponded to one of four dimensions, namely '*short'* and '*wide'* (related to the dish), and '*tall'* and '*thin'* (related to the glass).

Importantly, the gesture would either match the word (uttering '*short'* while gesturing to the dish's shortness), complement it (uttering '*short'* while gesturing to the dish's wideness), or mismatch it (uttering '*short'* while gesturing to the glass's tallness). A fourth condition (control) was created, where no gesture was present. In the gesture conditions, gesture onset would precede the word onset by 800 ms, and ERPs were taken to speech. As expected, both control and mismatching conditions led to a larger N400 effect compared to the matching condition, reflecting a greater ease of word processing for the latter case. Interestingly, there was no difference between the matching condition and the complementary condition, showing that when a gesture is semantically associated with the object described in the speech, it can be successfully integrated with the spoken entity, regardless of whether it is redundant or complementary. Thus, the study showed how gestures can enhance language comprehension, and how they can be integrated with speech at a higher semantic level (McNeill, 1992).

One of the most interesting and revealing aspects of these studies is that of the ability of iconic gestures to interact as a system of meaning with the meaning system of spoken words (Bernardis, Salillas & Caramelli, 2008). As Holler and Beattie (2003) put it, an iconic gesture and a spoken word will be jointly and harmoniously brought together to form a single and coherent linguistic mental representation. By doing so, the two channels, the visual and the auditory, provide the means for the addressee to obtain a far more concrete idea about the communicative and informative intention of the speaker (McNeill, 1992). Thus, iconic gestures constitute a valuable apparatus, allowing the comprehender to form a substantially more insightful and accurate representation, reflecting that of the speaker.

It becomes evident that, when it comes to comprehension, iconic gestures can help significantly with integrating the received information. When an iconic gesture does not fit with the line of information that comes from speech, it is expected to cause a substantial negative interference, resulting in a much greater effort in processing and deciphering the message as a whole. On the other hand, iconic gestures that are conceptually in line with the accompanied speech, can enhance the ease of processing, as highlighted in studies. Thus, iconic gestures are not mere hand movements that leave the addressee unconcerned, but rather highly meaningful messages that can affect to a considerable extent the task of semantic integration and can prove to be a highly valuable source of bolstering up the comprehension process.

In sum, a predictable context leads to the pre-activation of certain semantic features that are expected to appear in the sentence continuation, based on memory and prior experience. Moreover, iconic gestures temporally aligned and semantically congruent with the sentence continuation, have been shown to facilitate the process of semantic integration of this continuation. An interesting question that arises is how a predictable spoken discourse can interact with a meaningful iconic gesture, and to what extent this interplay between the two factors can lead to the semantic integration of a sentence-final target word with an even greater ease of processing, compared to the independent effect that these two factors have.

The present study

Evidence has accumulated that face-to-face interaction is a multidimensional process. Moreover, it appears that different lines of communication can be efficiently orchestrated and coordinated, such that, in a natural setting, we are inclined — either intrinsically or deliberately — to employ them as speakers, and to seek for them as addressees. In other words, the communicative process is multimodal and multifaceted, because it is simply more advantageous and gainful. The present study aimed to unravel to what extent iconic gestures and predictive processing combined can contribute to sentence comprehension. More specifically, the study sought to investigate how discourse-induced predictive processing can interact with signals distributed across different articulators, and to what extent this interaction can enhance sentence comprehension. But what exactly would such interaction mean?

As discussed earlier, prediction appears to be an exceptionally helpful mechanism. By utilizing the acquired experience, it is possible to infer valuable cues from the given context, in the form of reliable predictions about the possible continuation. Basically, this means that, during the unfolding discourse, certain semantic features are pre-activated. Consequently, entire mental representations of words are pre-activated and are readily available in the semantic memory, to be integrated into the developing discourse (DeLong et al., 2005). Eventually, if the context-induced predictive processing is accurate and the sentence continuation is consistent with the pre-activated features, the subsequently encountered signal (or signals) that the continuation contains will be more easily integrated. Now, as discussed, a spoken entity is shown to be more easily integrated when coupled with a semantically associated iconic gesture, compared to when encountered alone. As Holler and Levinson (2019) argue, this compound of elements from different articulators (gestalt) can facilitate semantic integration, despite the complexity that it entails. Thus, the present study claimed that, when certain mental representations are available in the working semantic memory due to a predictable context, integrating a prediction-consistent spoken entity that is merged with a semantically coupled co-speech iconic gesture will be even more efficient, (compared to when either of the two factors are omitted).

Exploring this question can provide additional elements that are missing from the current models of language processing that will, in turn, elucidate the unknown aspects of human communication. To this end, this study aimed to contribute to addressing these fundamental questions by focusing on the interaction of a predictable discourse and an iconic gesture, as manifested by the ease of processing of a sentence-final target word.

To address this question, an EEG experiment was conducted. The experiment focused on the N400 component as an index of semantic processing. Participants were exposed to audio-visual stimuli, while EEG was recorded. The auditory component of each trial consisted of two recorded sentences, featuring either a predictable or a non-predictable discourse context. With respect to the visual component of the trials, a recorded video was presented alongside the spoken sentence, with an actress producing a hand gesture. The auditory and visual components were temporally aligned, such that the stroke of the hand gesture was presented 130 ms before target word onset, thus presenting the iconic gesture and the target word almost simultaneously (see also Drijvers & Ozyurek, 2018). Critically, to assess the contribution of iconic gestures, it was imperative to contrast them with a control condition. Thus, the presented gesture constituted either a meaningful iconic gesture (iconic condition), or an irrelevant hand movement (control condition). Consequently, there were four conditions in total. The target words were concrete nouns (that can be gestured), each of which were rotated across the four conditions.

This design allowed me to examine the effect of the two factors, context and gesture, both combined, as well as independently. With regard to discourse context, the study investigated the impact of a predictable discourse context, as opposed to a non-predictable, with regard to the semantic processing effort of the sentence-final target word, as inferred by the N400 component. Importantly, the experiment was based on a between-participants design, meaning that a participant was presented to trials containing either predictable contexts or non-predictable contexts. To examine the effect of context only, the visually presented gesture was a control one. It was expected that there would be a main effect of discourse predictability. That is to say, compared to the non-predictable discourse condition, the predictable discourse would elicit a less negative N400 amplitude, thus reflecting facilitated processing of the target word.

To examine the main effect of gesture iconicity to the ease of processing of the target word, the discourse context was controlled, meaning that it was non-predictable, while the gesture was either iconic or a control one. Based on previous gesture studies, I hypothesized that, compared to the control gesture condition, iconic gestures would elicit a less negative N400 component. Such a result would indicate the facilitatory main effect that an iconic gesture can have, regarding the processing effort of its semantic affiliate, the target word.

With regard to the interaction of both factors, the present study sought to examine the interplay of discourse predictability and gesture iconicity, and its effect on language processing and the task of comprehension in particular. More specifically, the study investigated whether, and to what extent, a predictable context and an iconic gesture can interact, such that they will render the processing of the target word even more effortless, as reflected by the N400 amplitude. It was expected that the two factors will result into a significantly easier integration of the target word, relative to the other conditions. More specifically, I predicted that the presence of both factors would lead to a significantly reduced N400 amplitude relative to the conditions where only one of the two factors was present or both of them were absent, thus demonstrating the beneficial interplay of the two factors regarding the semantic integration of the sentence-final target word. Such a finding would indicate that, cost-effective linguistic entities which emerge from the fusion of elements from different articulators, can be integrated particularly more efficiently when relatable prediction-induced semantic features and representations have been pre-activated by the preceding discourse.

Method

Participants

Sixty-three healthy right-handed native speakers of Dutch (46 female) were recruited from the subject database of the Max Planck Institute for Psycholinguistics. No participant had hearing problems, neurological or developmental impairments, nor was diagnosed with dyslexia or any language deficit. Also, the participants' vision was normal or corrected to normal, and none of them was colour-blind. No participant had participated in any of the pre-tests. All participants provided written consent before taking part in the experiment. The study was approved by the Ethics Committee of the Social Sciences Faculty at the Radboud University and complied with the Declaration of Helsinki. Each participant signed written consent and was paid 18 Euro for participation. Due to too many excluded trials (above 25%, see Data pre-processing section for details) or too low overall accuracy on the comprehension question, three participants were excluded from all statistical analyses. Thus, the final set consisted of 60 participants (mean age = 24.3, range 18 – 34 years, SD = 3.47; 44 female).

Materials

Eighty concrete target nouns constituted the basis for the stimuli that were used in the experiment (mean Zipf frequency = 3.92, SD = 0.90, range = 2.06 – 6.47, Keuleers & Brysbaert, 2010; mean prevalence = 0.99, SD = 0.02, range = 0.91 - 1, Keuleers et al., 2015). These target words were presented in either a predictable or a non-predictable context. The context was always embedded in the first out of two Dutch spoken sentences. This first sentence was then followed by the second sentence, which ended with the target word. Together, the two sentences constituted a mini-story. Combined, the predictable and non-predictable contexts resulted in 160 mini-stories. For instance, for the target word '*kangaroo'* (in Dutch, '*kangoeroe'*), the mini-story with the predictable context was *'In Australia I visited local animals on a ranch. In a large outdoor pen I saw a kangaroo'*, whereas the mini-story embedded in the non-predictable context was '*All three looked at him intently. He approached cautiously the kangaroo'*.

Subsequently, each of the 160 mini-stories was coupled with an iconic gesture and a non-iconic (control) gesture, thus forming 320 unique stimuli in total. In the iconic gesture condition, the gesture was a conceptually meaningful hand movement, temporally aligned and semantically coupled with the target word. On the other hand, the control condition involved a meaningless hand movement, like scratching, again temporally aligned with the target word.

Regarding the control condition, the reason why a meaningless gesture was chosen over no gesture, lies in their difference in terms of processing levels. To elaborate, if for the control condition the total omission of a gesture was chosen, the spoken target word would be produced without an accompanying signal from a different articulator. However, by choosing a meaningless gesture as the control condition, the sentence-final signal that was created involved two different articulators, as with the iconic gesture condition. As Holler and Levinson (2019) point out, by binding signals from different articulators based on experience and without a hierarchical order, a so-called multiplex signal emerges at a perceptual pre-semantic level. However, as they argue, when fusing an iconic gesture and its spoken semantic affiliate, which are related with respect to meaning at the semantic level, the outcome is a gestalt (Holler & Levinson, 2019). In other words, a multiplex signal differs substantially from a gestalt, in terms of processing levels. Thus, this manipulation allowed me to compare the effect of elements bound at a higher semantic level of processing (iconic gesture condition), compared to elements fused at only a lower perceptual level (control gesture condition), with respect to semantic integration.

The creation of the audio-visual stimuli was done in the recording lab of the Max Planck Institute for Psycholinguistics. A female native speaker of Dutch was video-recorded as she uttered the mini-stories, while producing the iconic and the control gestures. Specifically, for the control gestures, the speaker would make meaningless hand movements, like scratching her chest or slightly pulling her blouse. The speaker’s clothes were neutrally coloured (dark black and blue). She stood in front of a unicolor curtain in the centre of the screen. While producing the spoken mini-stories, the speaker maintained a normal speaking rate, using a regular intonation.

At sentence onset, the speaker had her arms hanging in a casual way next to her body. In accordance with the instructions, she started producing the story and, close to the target word onset, she produced the gesture. For each stimulus, at least three versions were recorded, and then the optimum version was selected. The criteria for the selection were the consistency of speech and gesture across conditions, the level of naturalness of speech and gesture, as well as the recording’s quality (e.g., absence of video artifacts and no background noise).

Next, annotation was conducted on the onset and offset of certain events during the selected video recording, such as the gesture phrase, the gesture stroke, and the target noun. For the annotation, the software ELAN was used (Wittenburg et al., 2006). In addition, since the participants would have to pay attention to the hand gestures of the speaker, a mask was added to blur her face. Thus, no facial movements or expressions were visible to the participants. For the video editing, the software Adobe After Effects© was used. Subsequently, for each of the unique stimuli, the video was temporally situated with respect to the audio, such that the gesture stroke onset would precede the target word onset by 130 ms (see also Drijvers & Ozyurek, 2018).

On average, the 320 videos were 8302 ms long (SD = 1169, range = 6120 - 11952). On average, the occurrence of target word onset was after 6703 ms (SD = 1155, range = 4374 - 10336) into the recording. Importantly, this was comparable across predictable (M = 6715, SD = 1168, range = 4423 - 10336) and non-predictable conditions, as well as across iconic

(M = 6747, SD = 1159, range = 4374 - 10336) and non-iconic (M = 6660, SD = 1154, range = 4423 - 10010) gesture conditions.

Rating studies

To assess the cloze probability of the target words in the predictable and the non-predictable stimuli, two web-based sentence completion studies (Taylor, 1953) were conducted, using the software LimeSurvey GmbH. The first study concerned the 80 mini-stories including predictable contexts, whereas the second one concerned the other 80 mini-stories that included non-predictable contexts. 30 healthy native speakers of Dutch participated in the first study (22 female, M age = 26.2, SD = 3.5, range = 21 - 33) and 31 in the second study (18 female, M age = 24.0, SD = 4.0, range = 18 - 33). Importantly, none of the participants took part in the main experiment of the study.

In both studies, the participants' task was to carefully read the mini-stories up until (and including) the determiner that preceded the target word. After having read each mini-story, they were asked to fill in the word that, according to them, was the most likely to follow the determiner. A participant's response was coded as 'match' when the word in question was provided. On the other hand, if the response did not match the target word, the pairwise semantic distance between the response and the target word was calculated, using the Dutch version of Snaut (Mandera, Keuleers & Brysbaert, 2017). Subsequently, semantic distance values were converted into similarity values, by subtracting them from 1. The final step was to arrive at a cloze probability for each of the target words. This was done by summing up the 'match' responses (value of 1) and similarity values for the non-target responses (value between 0 and 1), and then divide the sum by the number of participants that had read the mini-story. For the predictable contexts, the average cloze probability was 0.85 (SD = 0.13, range = 0.51 - 1), whereas for the non-predictable contexts, the average cloze probability was

0.17 (SD = 0.07, range = 0.04 - 0.32). In conclusion, by successfully assessing the cloze probability of the target words in the predictable and the non-predictable stimuli, the rating studies reliably suggest that the classification of the mini-stories into those containing a predictable and those with a non-predictable context, was well established.

Next, a lab-based rating study was conducted, to assess the degree of compatibility between the iconic gestures and the target words. 32 healthy native speakers of Dutch participated in this study (23 female, M age = 23.0, SD = 2.9, range = 19 - 31). For this study, the software Presentation was used (Neurobehavioral Systems, Inc.). On each trial, the participants watched a video recording of each of the iconic gestures, without the accompanying audio. Then, they were instructed to provide a maximum of three guesses that best described the just-seen gesture. Finally, the participants were asked to provide a rating regarding the compatibility between the gesture and the target word it depicted. For the rating, a scale ranging from 1 (incompatible) to 7 (fully incompatible) was used. The probability of the target word being one of the three words given by the participant was on average 0.41 (SD = 0.32, range = 0 - 1), whereas the compatibility rating was on average 5.16 (SD = 1.24, range = 1.75 - 7). Thus, it was confirmed that the target words and the iconic gestures had a considerable and substantive degree of compatibility, as established by the rating study.

Experimental lists

For the main experiment, two target word lists were created, with respect to the discourse predictability condition (predictable or non-predictable condition). This means that one list was comprised of the mini-stories embedded in a predictable context, and one where the mini-stories had context that was non-predictable, thus forming the two groups of participants. Next, for each of these two lists, a second version was created, by shuffling the items across the two gesture type conditions (iconic and control). These two versions were balanced depending on the iconicity ratings. This means that the average rating, as well as the distribution of ratings, were the same for the iconic and control conditions in each of the two lists. For both pairs of versions (the pair with the predictable discourse and the pair with the non-predictable discourse), target words were coupled with a gesture, such that each target word was presented with the iconic gesture in one version and the control gesture in the other version. Thus, a total of 4 distinct stimulus lists were created.

Table 1. Descriptive Statistics for the 4 Stimulus lists are listed below. In first and second row, descriptive statistics about the iconic gesture ratings in versions 1 and 2 respectively are depicted. In the third row, descriptive statistics are presented for the cloze probability, regarding the predictable condition.

	Gesture Ratings		**Cloze Probability**
	Version 1	***Version 2***	
Valid	40	40	80
Mean	5.161	5.16	0.854
Std. Deviation	1.23	1.269	0.132
Variance	1.514	1.612	0.017
Range	4.44	5.25	0.491
Minimum	2.56	1.75	0.509
Maximum	7	7	1

For each of the 60 participants, stimulus lists were pseudorandomized with the use of the software Mix (van Casteren & Davis, 2006). In none of the subsequently created stimulus lists did the same iconicity condition repeat consecutively more than 3 times. Among the 60 participants, 30 were presented to a stimulus list version containing the predictable contexts, whereas the other 30 were exposed to a stimulus list version with the non-predictable contexts. Every stimulus list started by presenting two practice trials, followed by a comprehension question.

Procedure

The study was conducted at the Max Planck Institute for Psycholinguistics. Before the experiment, participants confirmed their good state of health, in line with the COVID 19-

related regulations. Then, they were informed about the content of the experiment, provided written consent, and filled in a short questionnaire, to ensure they were suitable for the experiment. Next, the preparation phase of the experiment took place, after which participants entered a sound-attenuating and electrically shielded experiment booth and were seated in a comfortable chair in front of a computer monitor, with two speakers placed on either side of it. Then, they received the instructions for the experiment. The stimuli were presented full screen on a 24-inch monitor, operating at a 1920 x 1200 native resolution. For the presentation of the audio-visual stimuli, the software Presentation was used (version 20.0, Neurobehavioral Systems, Inc.). Then, two practice trials were presented, followed by the 80 experimental trials, namely 40 featuring an iconic gesture and another 40 featuring a control gesture. To ensure that the participants were paying attention to the video, 20 of the trials were followed by a yes/no comprehension question (whether they just saw a red dot in the middle of the screen). These comprehension questions were spread throughout the experiment at irregular intervals. Every 20 trials, participants were given a self-timed break, before moving on to the next 20 trials.

EEG data recording

The EEG signal was recorded from a total of 32 electrodes, 27 of which (Fz, FCz, Cz, Pz, Oz, F7/8, F3/4, FC5/6, FC1/2, T7/8, C3/4, CP5/6, CP1/2, P7/8, P3/4, O1/2) were attached to the cap (AntiCap) and placed in accordance with the 10-20 convention. The EEG signal was recorded relative to the left mastoid (LM), which served as the online reference. One additional electrode was used to record activity from the right mastoid (RM) and four bipolar horizontal and vertical (right eye) electrooculogram (EOG) channels were used to monitor eye movements. The ground electrode was located on the Ground position (AFz). EEG signals were recorded using the BrainVision Recorder software (version 1.20.0401; Brain

Products GmbH), at a sampling rate of 1000 Hz, using a time constant of 8 s (0.02 Hz) and high cut-off of 100 Hz in the hardware filter.

Data pre-processing

Only the participants whose accuracy on the yes/no comprehension questions was 75% or higher, were included in the data pre-processing. This criterion led to the exclusion of... For the pre-processing of the raw data, the software BrainVision Analyzer (version 2.2.0.7383, Professional edition, Brain Products GmbH) was used. First, data were re-referenced to the average of the left and right mastoids channel. Then, data were filtered, by using a Butterworth IIR filter, with .01 Hz as a low cut-off and 30 Hz as a high cut-off. Next, the continuous data were segmented into epochs, ranging from -500 ms to 1000 ms, relative to the onset of the target word. Subsequently, ocular correction was applied, with the use of the Gratton and Coles (1983) algorithm. Using this algorithm, artifacts, such as blinks, were detected and corrected. Next step was to perform a semiautomatic artifact rejection. More specifically, BrainVision Analyzer highlighted the trials where channels values exceeded ±50 μV, and then these trials were examined and then kept or rejected, on an individual base. Importantly, after the artifact rejection step, at least 60 out of the initial 80 trials (75%) should have remained, for the participant’s data to be included in the statistical analysis. This criterion led to the exclusion of a total of 261 trials. Baseline correction was applied using a 200 ms window (-500 ms to -300 ms, relative to the onset of the target word), during which no gesture stroke had been presented yet.

From the 63 participants, two were excluded from the final analysis because of too few remaining trials (less than 75%, report accuracy for both participants) and 1 was excluded due to too many erroneous responses on the comprehension questions, thus leaving 60 participants for the analysis. With respect to the data of the remaining 60 participants, an average of 2.15 trials containing an iconic gesture were excluded per participant (SD = 2.58, range = 0 - 14), and 2.2 trials containing a control gesture were excluded per participant (SD = 2.67, range = 0 - 12). The average accuracy on the comprehension questions was 0. (SD = 0.02, range = 0.9 - 1).

Results

From the remaining 60 participants, and after having removed the excluded trials, a final total of 4539 trials was assembled. As described earlier, the experiment was designed such that half of the participants (30) were exposed exclusively to trials embedded in a non-predictable discourse, and the other half (30) were presented with trials containing a predictable discourse. Regarding the gesture element, for each participant, half of the trials (40) featured an iconic gesture, whereas the other half (40) featured a control gesture. The data were distributed across these 4 created conditions.

With regard to the 2 conditions featuring the non-predictable discourse, a total of 1124 trials (M = 37.47, SD = 3.17, range = 26 - 40) was collected for when a control gesture was presented, and a total of 1122 trials (M = 37.4, SD = 3.06, range = 28 - 40) for when the gesture was iconic. As for the other 2 conditions where the embedded discourse was predictable, a total of 1147 trials (M = 38.23, SD = 1.77, range = 34 - 40) was assembled when the gesture was a control one, and 1146 trials (M = 38.2, SD = 2.2, range = 31 - 40) for when an iconic gesture was presented.

Next, based on the data, grand-averaged ERPs were calculated for each of the 4 conditions. These are presented in Figures 1 and 2, regarding the 300 - 500 (main N400) and the later 600 - 800 ms time windows respectively, concerning the two representative electrodes, namely the Pz (placed on the parietal midline sagittal plane of the skull) and the Cz (placed on the central midline sagittal plane of the skull). In each of the two figures, the 4 conditions are represented with 4 different colours. The vertical dotted line represents the zero time-point which corresponds to the target word onset.

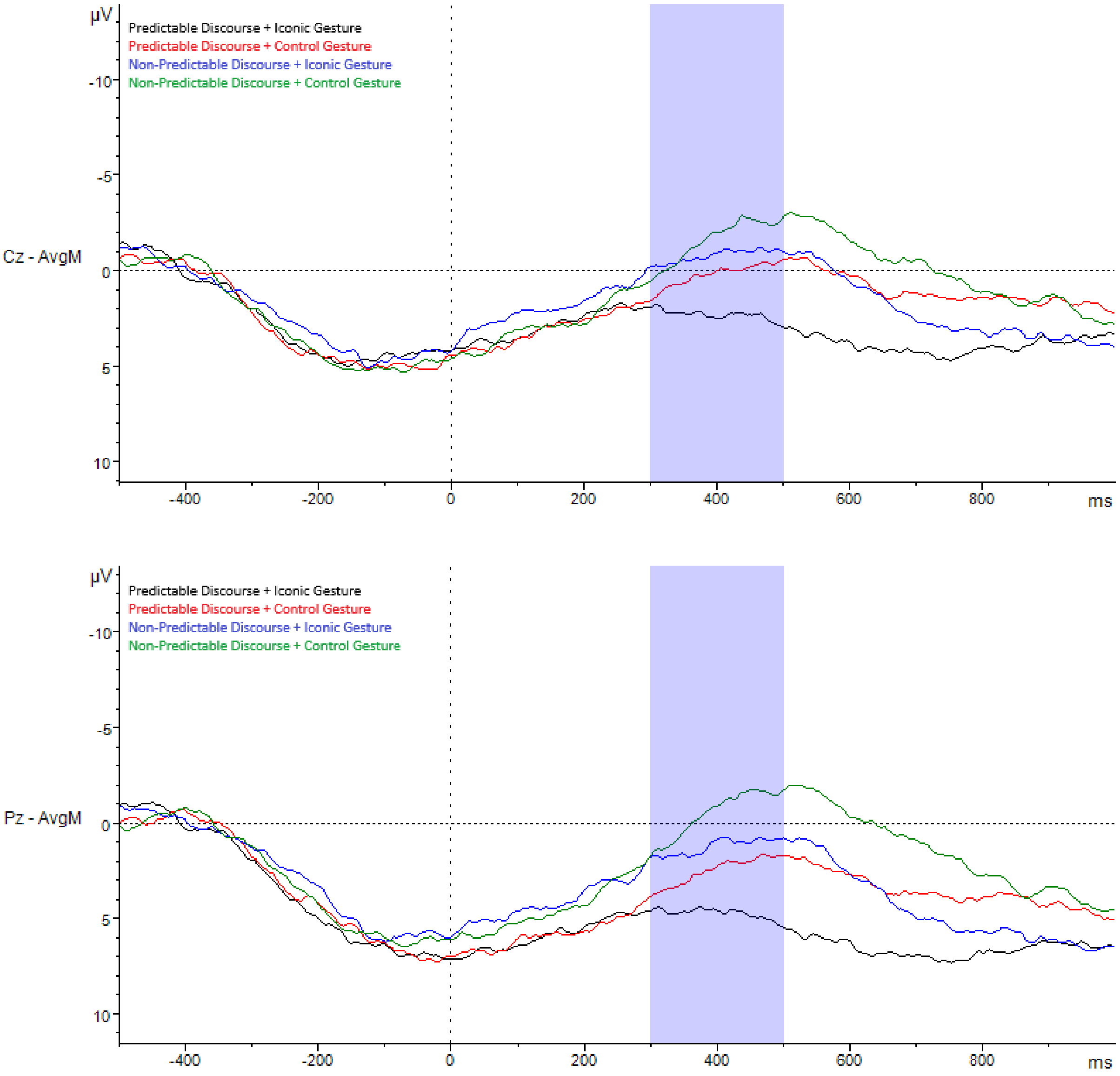

Fig. 1. 300 – 500 ms. (N400) time window. Grand-averaged ERPs for the two representative electrodes (i.e., the midline central [Cz] (top) and the midline parietal [Pz] (bottom)) are represented.

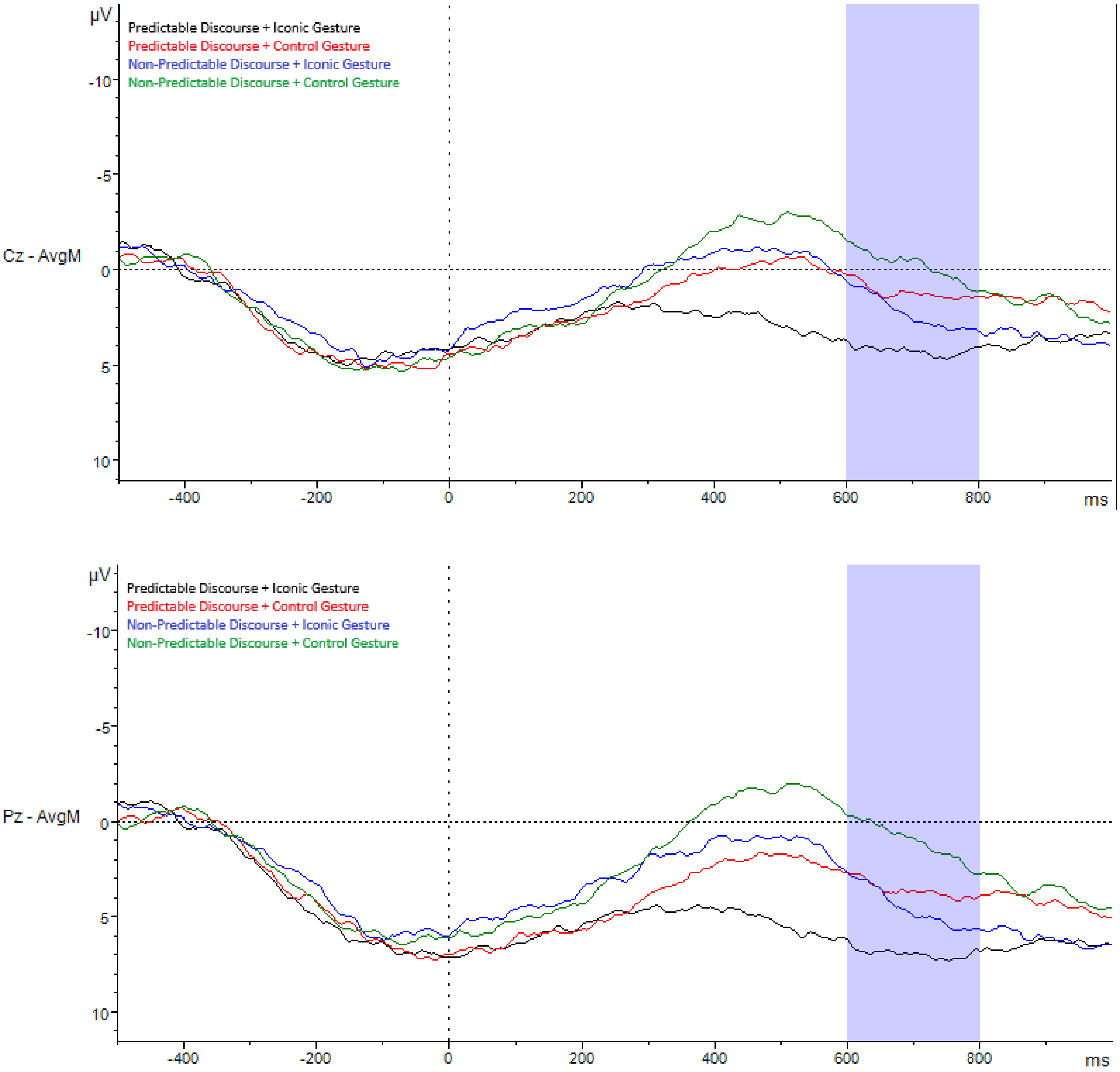

Fig. 2. 600 – 800 ms time window. Grand-averaged ERPs for two representative electrodes (i.e., the midline central [Cz] (top) and the midline parietal [Pz] (bottom)) are represented.

Based on previous research that has investigated the N400 component, it was determined to perform the statistical analysis on the time window of 300 - 500 ms after target word onset (Lau et al., 2008; Berkum et al., 1999; Kutas & Federmeier, 2011). On top of that, the visual inspection of the data suggested effects during a later window than the N400. Therefore, I ran an exploratory analysis on the subsequent time window of 600 - 800 ms after target word onset. For the statistical analysis, mixed-effects models were implemented, using the lme4 package (Bates et al., 2014) in R (R Core Team). The mixed-effects models were computed, with discourse predictability, gesture type, items and subjects as covariates. Discourse and gesture constituted the fixed factors, whereas items and subjects were the random factors. Since the N400 effect has been shown to have a parietal and central scalp distribution (Lau et al., 2008), we focused on the midline Pz and Cz electrodes, which are considered to reflect with substantial reliability and accuracy the effect of semantic integration, for both the 300 - 500 ms and the 600 - 800 ms time windows.

For the N400 component (300 - 500 ms time window), the formula that was used is the following: lmer(DV ~ Gesture * Discourse + (1+Gesture|Filename) + (1+Discourse|Item), data = data, control=bobyqa). As for the later ERP (600 - 800 ms time window), the R formula that was used is the following: lmer(DV ~ Gesture * Discourse + (1+Gesture|Filename) + (1+Discourse|Item), data = dataP, control=bobyqa). Also noteworthy is the fact that random effects had random slopes. The statistical analysis revealed a significant main effect of discourse predictability in the N400 time window (EST = 2.76274, SE = 0.91727, t = 3.012, $p < 0.01$). In the sequential time window of 600 - 800 ms after target word onset, by investigating the sustained negativity on the grand-averaged ERPs, a main effect of discourse predictability was again obtained (EST = 2.13134, SE = 0.95048, t = 2.242, $p < 0.05$).

Regarding gesture iconicity, a significant main effect was observed in the N400 time window (EST = 1.67583, SE = 0.54374, t = 3.082, $p < 0.01$). By looking at the next time window of 600 - 800 ms after target word onset, a main effect of gesture iconicity was again found (EST = 3.14463, SE = 0.53487, t = 5.879, $p < 0.001$) by the statistical analysis.

No evidence for an interaction of predictable discourse and gesture iconicity was observed. (N400 time window: EST = 0.76925, SE = 1.08748, t = 0.707, $p > 0.1$; 600 - 800 ms time window: EST = 0.19025, SE = 1.06974, t = 0.178, $p > 0.1$). All results are depicted below in Table 2.

Table 2. Estimates, std. error, t values and significance values for gesture iconicity, discourse predictability, and their interaction, with respect to their effect on the semantic processing of the target word.

Factors	*Estimate*	*Std. Error*	*t value*	*p value*
N400 (300 - 500 ms) time window				
Gesture Iconicity	1.67583	0.54374	3.082	< 0.01
Discourse Predictability	2.76274	0.91727	3.012	< 0.01
Gesture Iconicity & Discourse Predictability Interaction	0.76925	1.08748	0.707	> 0.1
600 - 800 ms time window				
Gesture Iconicity	3.14463	0.53487	5.879	< 0.001
Discourse Predictability	2.13134	0.95048	2.242	< 0.05
Gesture Iconicity & Discourse Predictability Interaction	0.19025	1.06974	0.178	> 0.1

Discussion

Prediction constitutes an essential and integral aspect of language comprehension. Evidence for this claim comes from a multitude of electrophysiological studies (Federmeier & Kutas, 1999; Szewczyk & Schriefers, 2013; Wicha et al., 2004; Van Berkum et al., 2005; DeLong et al., 2005; Pickering & Garrod, 2013). By focusing on the ERP of the N400 component, EEG studies have managed to demonstrate how the human mind engages in predictive processing. More specifically, it has been observed that, when embedded in a predictable context, a sentence-final target word leads to a reduced N400, compared to when the context is non-predictable. The interpretation is that the predictable context leads to a partial pre-activation of semantic features that relate to the ones of the target word, thus rendering its semantic processing more effortless.

Analogous research has been carried out with respect to the contribution of co-speech iconic gestures to the process of sentence comprehension. Many EEG studies have shown a facilitatory effect of iconic gestures regarding the ease of semantic integration of their semantic affiliates (McNeil, 1992; Kelly et al., 2004; Wu & Coulson, 2005; Bernardis, Salillas & Caramelli, 2008; Ozyurek et al., 2007). That is, a spoken entity, when coupled with an iconic gesture that is semantically associated with it, tends to be more easily integrated into a higher order representation. This has been demonstrated using variations in the N400 amplitude, which has been found to be less negative when a meaningful and semantically coupled iconic gesture is present, compared to a meaningless gesture or to the absence of such gesture all together.

Taken together, these studies suggest a language in-situ framework (Holler and Levinson, 2019), where language processing is enhanced and benefitted from a multimodal and multifaceted context. Apart from aiming to extend the notions of prediction and iconic gestures being facilitatory factors, the present study sought to investigate the interplay of these two factors, to find out whether their interaction can lead to an even greater facilitatory effect regarding sentence comprehension, and, if that is indeed the case, to what extent. In a between-participants design, half of the participants were presented with sentences embedded in a predictable context, whereas the other half were presented with sentences embedded in a non-predictable context. For each subject, half of the items contained an iconic gesture, whereas the other half contained a control one.

With regard to prediction, by focusing on the main time window, namely the N400 component, it was shown that there was indeed a main effect of discourse predictability (see Fig. 1). When the gesture was a control one, a predictable discourse did lead to a less negative N400, relative to a non-predictable context. This finding demonstrates how a sentence-final target word is processed with significantly greater ease, when the preceding predictable context has managed to instigate the pre-activation of semantic features that relate to the ones of the target word, compared to when this pre-activation has not taken place (see Bendixen et al., 2009; Wicha et al., 2004; Van Berkum et al., 2005). In other words, when the comprehender is initially presented to a predictable context, certain semantic elements that are linked to the upcoming target word are pre-activated. In turn, this causes the integration process of the target word to be significantly more effortless.

The facilitatory effect of discourse predictability becomes more evident when looking at the later time window of 600 - 800 ms after target word (see Fig. 2). As seen in the figure, the grand-averaged ERP of this later time window displays a significantly reduced amplitude of sustained negativity in the two conditions where sentences are embedded in a predictable context, relative to the two conditions where the context is non-predictable. It is worth noticing that a similar observation was made in the study by Rommers, Meyer, Praamstra and Huettig (2013). In their study, EEG was recorded, while participants were exposed to sentences containing a critical word (i.e., '*moon*') and a visually presented object preceding the acoustic onset by 500 ms, functioning as a potential predictable cue. Critically, the object either completely corresponded to the critical word (e.g., moon), or related to it with respect to its shape (e.g., tomato), or was entirely unrelated to it (e.g., rice). With respect to the EEG recording, the focus was on a time window spanning from 150-800 ms after critical word onset. Interestingly, the unrelated condition elicited a larger negative amplitude in both the 300 - 500 ms and the 500 - 700 ms time windows, compared to the related condition, demonstrating a prolonged negativity. This finding indicates that the process of integrating information in an unfolding discourse where the comprehender has been engaging in predictive processing, can potentially last longer than expected and exceed the traditional N400 time window of 300 - 500 ms.

As for gesture iconicity, iconic gestures had an enhancing effect, regarding the integration of the target word. As inferred from the N400 component, when discourse context was a non-predictable, an iconic gesture appeared to have a pronounced facilitatory effect. More specifically, the N400 amplitude was significantly reduced when an iconic gesture was present, compared to when the gesture was a meaningless one (control) (see Fig. 1). This main effect of gesture iconicity demonstrates the beneficial role that the visual channel can play when it is semantically compatible and aligned with the auditory channel. Consequently, when a word is coupled with a semantically corresponding iconic gesture, the combined information can be more easily integrated into a higher representational order.

As with the case of discourse predictability, the facilitatory effect of gesture iconicity can be further displayed when focusing on the subsequent time window of 600 - 800 ms after target word onset. As revealed by the statistical analysis, a main effect of gesture iconicity was inferred from the electrical activity in this time window. Overall, gesture iconicity led to the maintenance of a smaller amplitude of the ERP's negativity, when comparing to the control condition (see Fig. 2). This sustained negativity that was elicited can be interpreted as an additional index of conserved efficiency regarding the processing of the target word.

A relatable finding can be traced in the study by Wu and Coulson (2005). Their EEG study involved a mismatch paradigm, where both a main N450 component — comparable to the N400 of the present study —, and a late positive component (LPC), which was elicited around 740 ms after stimulus onset, were observed. Interestingly, regarding the LPC component, a larger positivity was observed for congruous items. This means that the incongruous items elicited a more negative amplitude in this particular time window, indicating that, as with the present study, an extended negativity was observed when semantic processing was more effortful. Thus, overall, when looking at both time windows in the present study, namely the main N400 (300 - 500 ms) and the later (600 - 800 ms), there appears to be a prolonged main effect of gesture iconicity, as inferred by the negativity amplitude across conditions, thus highlighting the beneficiary effect that iconic gestures can have on the process of integrating the target word.

A particularly interesting aspect regarding gesture iconicity can be traced in the non-predictable discourse condition. It appears that, even when the comprehender is presented with a non-predictable context, an iconic gesture coupled with its spoken semantic affiliate can make the difference with respect to the its semantic integration, compared to when the gesture is non-iconic. This finding indicates that, even without having managed to pre-activate a range of certain semantic features, the provision of a gestalt is highly valuable and more helpful that a signal deriving from a single articulator. A gestalt, formed by signals from different articulators which are bound together at both a lower perceptual and a higher semantic level, can trigger a parsing mechanism that is able to integrate information remarkably efficiently, despite the complexity that it entails. Thus, the compound of speech and an iconic gesture is arguably an adequate communicative instrument on its own.

It has been established that both discourse predictability and gesture iconicity had a main effect with respect to the processing of the target word. These two factors affected the integration process individually, as shown by the statistical analysis. The main goal of the present study, however, was to investigate the interaction between discourse predictability and gesture iconicity. The statistical analysis revealed that the interaction effect was not significant (see Table 2). That is, although a predictable context combined with an iconic gesture rendered the processing of the target word more effortless compared to the other three conditions, the difference was not statistically significant.

The fact that no significant effect was yielded from the interaction of discourse predictability and gesture iconicity can have several interpretations. One the one hand, it could be the case that, when certain mental representations have been pre-activated due to a predictable context, the presence of an iconic gesture does not significantly enhance the integration process. This would imply that the summoning of prediction-induced semantic features is sufficient, to the extent that the subsequent integration of a prediction-consistent spoken entity does not require a co-speech iconic gesture to be more efficient. In case this interpretation holds true, it would raise interesting questions about the potential of gestalts. Further research might be necessary to unravel this important question.

On the other hand, there is another possible explanation for the absence of an interaction effect. As demonstrated from the main effect of gesture iconicity, when encountering a spoken entity coupled with an iconic gesture, its processing effort is significantly reduced, compared to when the iconic gesture is omitted. Thus, the fact that no interaction was found might indicate that the pre-activation of prediction-induced semantic features does not necessarily 'pave the way' to especially favour the integration process when the spoken entity is coupled with an iconic gesture, compared to when it is not. Of course, this interpretation requires further investigation as well, to examine the potential of predictive processing with respect to multimodal language processing.

There is also a possibility that the absence of an interaction effect might be connected to the between-subjects design that was implemented, meaning that the design allowed for a substantial variability between participants to occur, which can be attributable to several reasons. For instance, one factor that could conduce to this variability concerns the quality of the participants' lexical representations. Lexical representations owe their acquisition and development to the amount of encounter and exposure to the corresponding words throughout time, and the course of this progress affects the quality of each lexical representation. In turn, having a poor representation of a certain lexical entity could potentially negatively affect the comprehension process of a corresponding word— the target word in the case of the present study — when encountered, and render it more effortful, compared to someone who has a richer lexical representation of the same entity. In turn, this variability between participants essentially translates into variability in the discourse conditions for each participant, which is highly likely to have contributed to the lack of interaction between discourse predictability and gesture iconicity. On the other hand, it can be argued that, by running mixed models, it was made possible to control for such an effect of the between-participants variability. Future research might be needed to further address this issue.

To conclude, the present study demonstrated the significant facilitatory effects that discourse predictability and gesture iconicity can have on language comprehension, as displayed by their independent effects. In addition, as inferred by the examined ERPs, the mutual presence of a predictable discourse and an iconic gesture did not result into a significantly more effortless processing of a target word. This finding indicates that, although the interaction appears to have an advantageous impact, the limited extent to which it does poses some interesting questions regarding the gains of a multimodal and multifaceted communicative context. Yet, the study's findings do not call for the rejection of the in-situ multimodal framework.

Literature

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in cognitive sciences*, *11*(7), 280-289.

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235-1243.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience*, *29*(26), 8447-8451.

Berkum, J. J. V., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of cognitive neuroscience, 11*(6), 657-671.

Bernardis, P., Salillas, E., & Caramelli, N. (2008). Behavioural and neurophysiological evidence of semantic interaction between iconic gestures and words. *Cognitive Neuropsychology, 25*(7-8), 1114-1128.

Boudewyn, M. A. (2015). Individual differences in language processing: Electrophysiological approaches. *Language and Linguistics Compass, 9*(10), 406-419.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, *8*(8), 1117-1121.

Drijvers, L., & Özyürek, A. (2018). Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions. *Brain and language*, *177*, 7-17.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of memory and Language*, *41*(4), 469-495.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology*, *55*(4), 468-484.

Holler, J., & Beattie, G. (2003). How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process?

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639-652.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and language*, *89*(1), 253-260.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, *42*(3), 627-633.

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology*, *68*(8), 1665-1692.

Kropotov, J. D. (2016). *Functional neuromarkers for psychiatry: Applications for diagnosis and treatment*. Academic Press.

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.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature reviews neuroscience*, *9*(12), 920-933.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57-78.

McNeill, D. (1985). So you think gestures are nonverbal?. *Psychological review*, *92*(3), 350.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

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.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences*, *36*(4), 329-347.

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, *51*(3), 437-447.

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, *68*(4), 297-314.

Taylor, W. L. (1953). “Cloze procedure”: A new tool for measuring readability. *Journalism quarterly*, *30*(4), 415-433.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443.

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior research methods*, *38*(4), 584-589.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of cognitive neuroscience*, *16*(7), 1272-1288.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556-1559).

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

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and language*, *101*(3), 234-245.