Cognition 236 (2023) 105449

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

# When a look is enough: Neurophysiological correlates of referential speaker gaze in situated comprehension☆

Torsten Kai Jachmann *, Heiner Drenhaus, Maria Staudte, Matthew W. Crocker

*Language Science and Technology, Campus C7, Saarland University, D-66123 Saarbrücken, Germany*

ARTICLE INFO

Dataset link: https://osf.io/uzktp/

*Keywords:*
ERP
Speaker gaze
Referential expectation
PMN
N400
P300
P600

ABSTRACT

Behavioral studies have shown that speaker gaze to objects in a co-present scene can influence listeners' expectations about how the utterance will unfold. These findings have recently been supported by ERP studies that linked the underlying mechanisms of the integration of speaker gaze with an utterance meaning representation to multiple ERP components. This leads to the question, however, as to whether speaker gaze should be considered part of the communicative signal itself, such that the referential information conveyed by gaze can help listeners not only form expectations but also to confirm referential expectations induced by the prior linguistic context. In the current study, we investigated this question by conducting an ERP experiment (N=24, Age:[19,31]), in which referential expectations were established by linguistic context together with several depicted objects in the scene. Those expectations then could be confirmed by subsequent speaker gaze that preceded the referential expression. Participants were presented with a centrally positioned face performing gaze actions aligned to utterances comparing two out of three displayed objects, with the task to judge whether the sentence was true given the provided scene. We manipulated the gaze cue to be either Present (toward the subsequently named object) or Absent preceding contextually Expected or Unexpected referring nouns. The results provided strong evidence for gaze as being treated as an integral part of the communicative signal: While in the absence of gaze, effects of phonological verification (PMN), word meaning retrieval (N400) and sentence meaning integration/evaluation (P600) were found on the unexpected noun, in the presence of gaze effects of retrieval (N400) and integration/evaluation (P300) were solely found in response to the pre-referent gaze cue when it was directed toward the unexpected referent with attenuated effects on the following referring noun.

## 1. Introduction

In situated communication, speaker gaze to objects they are about to mention is closely synchronized with the speech signal (Griffin & Bock, 2000; Kreysa, 2009; Meyer et al., 1998), offering listeners a reliable cue that has been shown to facilitate communication in a variety of settings and behavioral measures (Argyle & Cook, 1976; Hanna & Brennan, 2007; Staudte & Crocker, 2011; Staudte et al., 2014). This raises the question of whether speaker gaze is simply some "non-linguistic" cue that listeners learn to take advantage of, or whether it should rather be considered part of the communicative signal itself. To address this question, we examine the extent to which neurophysiological responses to referential gaze, prior to hearing the corresponding referring expression itself, are similar to those normally elicited only once the referring expression is actually heard.

Gaze has been shown to be an important social cue that is recognized and utilized, even by infants, to establish joint attention, gaze following behavior, and the assignment of mental attribution (Baron-Cohen, 1997; Butterworth & Jarrett, 1991; Mundy & Gomes, 1998; Symons et al., 1998). Further it was found that, in adults, the gaze to an object influences its perception as for example an attributed graspability of an object as shown in an fMRI study by Becchio et al. (2008) where the activation of brain regions involved in hand-object coordination were recorded when participants simply observed gaze to a graspable object. Studies by Driver IV et al. (1999) and Friesen and Kingstone (1998) have provided evidence that gaze elicits reflexive orientation in observers, that goes beyond volitional control, even when gaze is uninformative or misleading. While Uono et al. (2014) have also shown that those aspects of gaze can also be induced by other cues

☆ Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) — Projektnummer 232722074 — SFB 1102/Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — Project-ID 232722074 — SFB/CRC 1102.
* Correspondence to: Universität des Saarlandes, Fachrichtung Sprachwissenschaft und Sprachtechnologie, Geb. C7.1, Raum 1.16, Campus D-66123, Saarbrücken, Germany
*E-mail address:* jachmann@lst.uni-saarland.de (T.K. Jachmann).

https://doi.org/10.1016/j.cognition.2023.105449
Received 26 August 2020; Received in revised form 16 March 2023; Accepted 23 March 2023
Available online 6 April 2023

such as arrows, the underlying mechanisms involved in the perception of gaze and (e.g.) arrows seem to differ (Hietanen et al., 2008; Ristic et al., 2002). It stands to reason that gaze holds a special role amongst situational communication cues based on its ubiquitous availability and in that its reflexive use potentially yields a more relatable and consistent index of the gazer's mental processing than consciously utilized cues.

Various studies on the influence of speaker gaze on listeners' sentence comprehension have provided evidence for the view that this non-verbal cue conveys the speaker's referential intention and is interpreted as part of the communicative signal (Hanna & Brennan, 2007; Jachmann et al., 2019; Staudte & Crocker, 2011; Staudte et al., 2014). A study by Argyle and Cook (1976)has shown that interlocutors are aware of each other's gaze and that they direct their gaze toward objects under discussion. The establishment of joint attention through gaze has been found in various primates (Thomsen, 1974; Tomasello et al., 1998) and reported to be a supporting influence in language acquisition (Tomasello & Farrar, 1986; Tomasello & Kruger, 1992). Gaze has further been reported to express the focus of a speaker's attention and can express the preparation of referring expressions or the allocation of relevance to the current discourse (Griffin & Bock, 2000; Hanna & Brennan, 2007). More precisely, it has been found that speakers gaze at an object about 800–1000 ms prior to its mention (Griffin & Bock, 2000; Kreysa, 2009; Meyer et al., 1998). Similarly, listeners have been shown to direct their gaze toward objects about 200–400 ms after hearing the onset of the referring expression (Allopenna et al., 1998; Tanenhaus et al., 1995). Eye-tracking studies have further provided evidence that listeners direct their gaze toward anticipated targets even before their mentioning if the sentence provided strong support for a specific target over others, as for example shown in a study by Altmann and Kamide (1999). In their study, listeners were provided with a scene in which a verb (e.g., "eat") fit one displayed object (e.g., "cake") but none of the other displayed objects. Upon hearing only the constraining verb, participants directed their gaze toward the anticipated target. This was not the case for verbs that were not biased to a particular target (e.g., "move"). This shows that human gaze composes a special, extremely informative signal. There is evidence however, that schematic gaze also elicits similar effects to human gaze. Chen and Yeh (2012) replicated the effect of direct gaze facilitating the detection of faces as compared to averted gaze (Stein et al., 2011) for schematic faces, even in conditions where no facial features beside eyes were present. Farroni et al. (2004) have shown that even infants show similar gaze following behavior to schematic gaze depiction as to photographs of faces, such that they follow a directional gaze after a mutual gaze in both cases. In a study investigating the effect of initial mutual gaze on gaze following behavior, Böckler et al. (2011) found congruency effects of gaze only when participants first observed a mutual gaze between two agents. Importantly, the effects were similar for both real faces as compared to schematic depictions of faces. We therefore build on Jachmann et al. (2019) who also employed stylized gaze cues to study the role of human-like speaker gaze for building expectations.

Building on the behavioral studies by Staudte and colleagues (2010, 2011, 2014), Jachmann et al. (2019) conducted an Event-Related Potential (ERP) study that provided robust evidence that speaker gaze was utilized by listeners to form expectations about the unfolding sentence, facilitating sentence comprehension when the gaze to an object was reliable (congruent with the subsequently named referent) and, in turn, impairing the referring noun's processing when preceding gaze was unreliable (incongruently directed toward an object other than the subsequently named referent).

In two experiments (Jachmann et al., 2019), participants were presented with a stylized face performing gaze actions toward objects positioned around it. Those gaze actions were aligned with an auditory utterance presented to the participants that compared two of the three present objects with one-another. The presented sentences were of the form

(1)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Verglichen* | *<Gaze1>* | *mit* | *dem* | *Auto,* | *<Gaze2>* | *ist* | *das* |
| Compared | <Gaze1> | to | the | car,<br>NP1 | <Gaze2> | | the |

| | | | |
|---|---|---|---|
| *Haus* | *verhältnismäßig* | *groß,* | *denke ich.* |
| house<br>NP2 | is relatively | big,<br>Comp | I think. |

Gaze cues were presented 800 ms before the onset of each referring noun. The structure of the sentence (NP1–NP2–Comparative) did not provide any preference for either of the objects to be named as only the comparative at the end of the sentence (groß/big) put the presented objects in a relation. In the experimental conditions, the first gaze cue (preceding NP1) was always congruent toward the subsequently named object (the car in the example), while the gaze cue preceding the second noun (NP2) *'house'* was manipulated to be either Congruent toward the subsequently named object, Incongruent toward the object unmentioned throughout the course of the sentence, or Uninformative. Uninformative gaze was represented in two ways: either an Averted gaze cue, toward an empty position (in their first experiment), or a Mutual gaze cue, redirected toward the listener (in their second experiment).

The results of those experiments clearly indicated that listeners utilized the gaze cues to form expectations about the unfolding sentence, as revealed by three distinct ERP components time-locked to the onset of the target noun (NP2). More specifically, an increased Phonological Mismatch Negativity (PMN) response in the time-window between 150 and 300 ms indexing a phonological matching mechanism was only found when the perceived phoneme of the spoken noun did not match the expected word form (Incongruent condition) or when the phoneme could be utilized to select the actual referent when no prior expectations about the referent were possible (Uninformative conditions). More broadly speaking, the PMN appeared to be related to the amount of information the phoneme provided, which is in line with a proposed interpretation of the PMN by Hagoort and Brown (2000) (see also Connolly & Phillips, 1994). Further, an increased N400 amplitude between 300 and 450 ms was present in the Incongruent and Uninformative conditions. This effect was interpreted in line with the Retrieval Integration Account (RIA) (Brouwer et al., 2012), in which the N400 is interpreted as indexing a word meaning retrieval mechanism (Federmeier & Kutas, 1999; Kutas & Federmeier, 2000). The increased N400 amplitude expresses the retrieval effort of a word that was either not anticipated (Incongruent condition) or for which no prior expectations were formed (Uninformative conditions). Also in line with the RIA was the P600 component between 600 and 800 ms indexing a mechanism of integration with sentence meaning where an increased P600 response only for Incongruent gaze expresses the necessity to revise the mental representation formed on the basis of the prior context, including the gaze cue.

In the reported study, Jachmann et al. (2019) proposed a Situational Integration Account (SIA) in which not only the utilization of linguistic content is used to form a mental representation but that this representation of the situation is built and extended based on any relevant information in situated communication, including non-linguistic cues such as (potentially) gestures and, specifically, gaze (additional evidence for the use of such dynamic visual cues can be found in Kreysa et al. (2018)). As such, the proposed account extends the situation model as described by Zwaan and Radvansky (1998). Where their characterization of a situation model is mostly linguistic in nature, while also drawing from world knowledge to frame the input, we propose that a situation model incorporates further information. Specifically, while they argue that a reader creates a situation model when being exposed to the first sentence/information describing a situation, we suggest that in situated face-to-face interactions, interlocutors are exposed to more information beyond the purely linguistic information. As such, we propose that the situation model includes any discourse relevant information, including linguistic information, but also visual and potentially environmental information, such as the location of the interaction, the present objects and people in the vicinity, as well as the interlocutors facial expressions and gaze.

Thus, the situation model according to the proposed SIA is not initially constructed when a recipient is exposed to linguistic information (e.g., speech) but well before that when being exposed to the setting surrounding the linguistic information (e.g., location, objects in the vicinity, etc.) and changes or interactions within that scene (such as gaze to objects). The former is also supported by findings from Tanenhaus et al. (1995) and Trueswell et al. (1999): Both studies demonstrated how the visual scene was utilized to resolve syntactic ambiguities in instructive sentences. In their experiment, they presented participants with objects (e.g. frog dolls) and locations (e.g. napkins, boxes) with the task to move objects from one position to another. The task was provided via a verbal instruction of the form “Put the frog on the napkin in the box”. Here, the phrase “on the napkin” leads to a local ambiguity as it could either be parsed as a modifier of “frog” or a destination for the frog doll to be put at. They demonstrated that participants would interpret the phrase differently, depending on competitors in the scene: When only one frog (positioned on a napkin) was present, the phrase was interpreted as the frog doll’s destination (an additional empty napkin). However, in the presence of another frog doll that was not positioned on a napkin, the phrase was interpreted as a modifier of frog, disambiguating the two potential targets of the action. These shifts of interpretation of the same phrase were evident from eye-tracking data, which showed that participants looked at the frog on the napkin in the ambiguous condition on the onset of “napkin” while they looked more toward the empty napkin in the unambiguous condition. This shows that the information provided in the scene is rapidly utilized to inform sentence interpretation (see also Kamide et al., 2003; Knoeferle & Crocker, 2006).

The SIA posits further that the visual scene is not only utilized for disambiguation purposes in scenarios that require direct interaction with the scene, which are evident from gaze actions in response to linguistic content (e.g. the naming of an object and/or a location) but also guides linguistic expectation. In the frog example mentioned above, this would entail that, after the identification of the target object, participants would also anticipate the upcoming target destination (e.g. the other empty napkin or box). Further, the SIA assumes that additional information in the scene, such as gaze to one of those possible destinations, would then narrow those expectations down to the gazed-at destination, leading to expectations for a specific lexical continuation. Importantly, while gaze is part of the visual environment, it is qualitatively different from the objects and events in the visual scene in that it provides a dynamic indicator of the speakers attention, including referential intentions that are synchronized with the speech signal. As such, we hypothesize that this influence of gaze on incremental and anticipatory comprehension processes is more appropriately considered as part of the communicative signal rather than as just another component of situational context that the signal may make reference to.

Further, the situation model and updating process described by Zwaan and Radvansky (1998) differs slightly from the updating process proposed under the SIA. They propose a sentence by sentence (or clause by clause) situational model that is created based on the information obtained from the current sentence and integrated with the previous situation model to form a coherent representation. We propose that the updating process is more incremental than that and does not require the construction of an additional model which is integrated as a whole, but that new information is utilized as soon as it is available to update the situation model. Finally, the situation model as described by Zwaan and Radvansky (1998) is primarily put into context with language comprehension and memory. Under the SIA, we further propose the situation model to be part of interactions, such that the information contained in the situation model is monitored and utilized to formulate task-relevant responses (e.g., verbal responses, button presses, motor

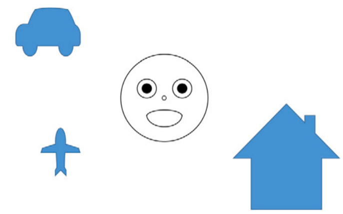

**Fig. 1.** Example Screen as presented to Participants.

actions, etc.). Thus, under the SIA the situation in the mental representation consists of various (potentially) communication relevant elements. These include (but may not be limited to) the thus far created context, communication relevant entities in the vicinity (or memory) of the interlocutors, gaze and potentially other gestures. The mental representation of the situation is continuously constructed, adjusted, and extended in order to ease linguistic interaction by the ability to anticipate upcoming elements of the communication.

In sum, the results from Jachmann et al. (2019) show that speaker gaze is an integral part of the communicative signal, used by listeners to form expectations which can facilitate both retrieval and integration of subsequent linguistic content. To the extent that gaze is indeed a component of the communicative signal and not just considered part of the context, this predicts that referential gaze itself toward a contextually unexpected referent may convey sufficient referential information to elicit a similar ERP signature as the referring noun itself. This is an open question since gaze arguably is not a direct substitution of a noun, but instead typically accompanies referential expressions (Sekicki & Staudte, 2018).

The present study investigates this question and further tests the assumptions made under the SIA, which are that any sort of information — here, visual and linguistic — is utilized to inform the mental representation of the situation, incrementally. We conducted an ERP experiment in which the role of gaze and linguistic content was reversed compared to that of Jachmann et al. (2019), such that the linguistic content of a sentence would establish referential expectations about the unfolding sentence. The gaze cue that preceded the naming of the gazed-at object could then in principle be utilized to confirm or falsify those expectations before the actual referent was named. In order to achieve this, we changed the sentences from the structure as presented above to an NP1–Comparative–NP2 structure, which shifts the point at which referential expectations can be derived from the gaze cue (Jachmann et al., 2019) to the comparative (the present study). In more detail, example sentence (1) above was changed to

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| (2) | *<Gaze1>* | *Das* | *Auto* | *ist* | *kleiner* | *als* | *das* | *<Gaze2>* |
| | <Gaze1> | The | car<br>NP1 | is | smaller<br>Comp | than | the | <Gaze2> |
| | *dargestellte* | *Haus,* | *denke ich.* | | | | | |
| | displayed | house,<br>NP2 | I think. | | | | | |

While both sentences (examples 1 and 2) have the same semantic content, expectations about the second noun phrase can be made based on the linguistic content as soon as hearing *’kleiner’/’smaller’*, when presented with the same scene as utilized in the second experiment by Jachmann et al. (2019), depicted in the time line presented in Fig. 1.

The referent would either be Expected given the preceding linguistic content (the referent that would lead to a correct utterance in relation to the provided scene, as the *house* in example (2) in relation to Fig. 1) or Unexpected (the referent that would lead to a false statement, as the *airplane* in example (2) in relation to Fig. 1). In order to test whether gaze is not only used to form expectations but to also confirm them, we manipulated the occurrence of the gaze cue preceding the naming of the second referent in the sentence so that it was either

Present or Absent. If gaze was Present it was always directed toward the object subsequently named. Broadly, we hypothesized that Unexpected continuations would lead to similar effects as reported in Jachmann et al. (2019), indexing the involvement of a phonological matching mechanism (PMN), a word meaning retrieval mechanism (N400) and a mechanism of integration with sentence meaning (P600) with increased responses for each of the indexing components when the referent was Unexpected. In the absence of gaze, we predicted the same effects as reported in Jachmann et al. (2019) on the noun, such that each of the three predicted components (PMN, N400 and P600) was expected to express increased responses to Unexpected referents. Further, we hypothesized that — to the extent that speaker gaze would be perceived as part of the communicative signal — the effects of expectancy evaluation would instead be found on the gaze cue if present, and possibly with attenuated responses in the following noun region since the noun provided no new information when gaze was present. Predictions regarding the PMN on the gaze cue are less clear, given that gaze does not contain phonological content: Given that there is an expectation for a specific word prior to the gaze cue, it could be that comprehenders retrieve the phonology of the gazed at object and match it with the expectations. On the other hand, the PMN as reviewed above has, to our knowledge, thus far only been elicited in response to the acoustic signal, in which case no PMN on the gaze cue would be expected.

## 2. Methods

As in Jachmann et al. (2019), we utilized a stylized representation for the face and objects to keep the cognitive demand of recognition of the presented objects as well as the face as low as possible. This also includes eliminating the influence of other facial cues, such as head movement or facial expressions, which have been shown to influence the perception of gazed at objects (Bayliss et al., 2007, 2006). Further, as the face was paired with a synthesized voice in order to maintain control over the prosodic pattern, co-articulations, etc., the stylized depiction aimed to support the connection between the face as the speaker of the sentence and the voice. Similar stylized faces have been used in various studies investigating the effects of gaze (Friesen et al., 2004; Kuhn & Kingstone, 2009) and provided evidence that actual faces and stylized faces elicit similar effects (e.g., Böckler et al., 2011). Most relevantly, Jachmann et al. (2019) found a robust influence of such stylized gaze cues in ERP components associated with phonological processing, lexical retrieval, and semantic integration.

The present experiment, while utilizing the same visual displays as Jachmann et al. (2019), employed a different sentence structure in which the linguistic content could be used to form expectations about the second referent prior to encountering the referring expression (NP2) — including gaze, if present. The gaze cue thus could be used to confirm or falsify those expectations, but was neither required to understand the sentence nor to anticipate the continuation of the sentence. The stimuli were created using the same objects and screen arrangements as in the second experiment of Jachmann et al. (2019) with a diagonal placement of the objects around the face (for comparison, see Fig. 1). A summary of all used objects can be found in Table 1. The CereVoice TTS system's Alex voice (Version 3.2.0) was used to synthesize the sentences that were presented to the participants.

In order to create a scenario in which the sentences support the forming of expectations based on the linguistic content, the sentences were created in a **Comparative-Gaze-Referring Noun** structure as in example 2 above.

It should be noted that, while each sentence contains two gaze cues toward objects as well as two referring nouns, the terms "referent" and "referring noun" throughout this paper will only refer to the second referent as manipulations only took place in that region. Similarly, "gaze" will refer to Gaze2.

In contrast to the previous experiments, the linguistic content of the sentence establishes a clear expectation for an upcoming referent as soon as the comparative. Together with the co-present visual scene, the sentence onset (*'the car is smaller than'* in example (2) above) supports expectations for *'house'* to be the upcoming referent, as it is the object that is bigger than the car. This expectation can in principle be formed prior to both the gaze toward the object and its subsequent naming. With this change, the role of the comparative also shifts: In the two experiments of Jachmann et al. (2019), the clause final comparative was only useful to determine whether the sentence was true or false, whereas in the present study the comparative serves as predictive linguistic information to anticipate the upcoming referent (cf. Staudte & Crocker, 2011).

**Table 1**
Summary of the objects presented in the experiment with their English translation separated by grammatical gender.

| Masculine | Feminine | Neuter |
|---|---|---|
| Baum (tree) | Blume (flower) | Auto (car) |
| Blitz (bolt) | Brezel (pretzel) | Blatt (leaf) |
| Fisch (fish) | Gießkanne (watering can) | Boot (boat) |
| Handschuh (glove) | Hand (hand) | Flugzeug (airplane) |
| Hut (hat) | Lampe (lamp) | Haus (house) |
| Stern (star) | Maske (mask) | Kreuz (cross) |
| Stiefel (boot) | Tasche (bag) | Rad (wheel) |
| Tisch (table) | Wolke (cloud) | T-Shirt (t-shirt) |

The gaze cue preceding the mention of the second object was manipulated such that it was either Present or Absent. If, however, the gaze cue was present, it was always directed toward the subsequently named referent.

Overall, this manipulation led to a 2 × 2 design with the factors Gaze Presence (Present/Absent) and Linguistic Expectancy of the referent in respect to the provided scene (Expected/Unexpected). It should be noted that, in this experimental design, expectancy and the truth value of the sentence are directly related. A referent is expected when its mentioning would make the sentence true. Hence, unexpected continuations by design lead to false statements. In the previously mentioned example related to the scene depicted in Fig. 1, this would mean that such a false statement was

(3) *<Gaze1> Das Auto ist größer als das <Gaze2>*
<Gaze1> The car is bigger than the <Gaze2>
*abgebildete Haus, denke ich.*
displayed house, I think.

The time line depicted in Fig. 2 shows the course of an experimental trial: At the beginning of each trial, participants were presented with three objects. After 1000 ms, additionally a face appeared in the center of the screen from which point on participants were instructed to focus their gaze on the nose of the face to reduce eye-movement artifacts. After another 1000 ms, the gaze of the presented face shifted toward the first referent in the sentence (Gaze1). Due to the position of the first referring expression ('car' in Example 2) at the beginning of the sentence, this gaze shift preceded the onset of the sentence in order to maintain the 800 ms difference between the fixation of a referent and its mentioning. In experimental trials, the first referent in the sentence was always the medium sized object. For both other objects in the scene, the comparative would already suffice to conclude whether the sentence was true or false. For example, if the first referent was the largest object on the screen, hearing 'smaller' would render the sentence false as there is no larger object present (and vice versa for the smallest object). At sentence onset, the face retained its gaze on the first referent and opened the mouth in order to reinforce the impression of the face as being the speaker of the sentence. Following the mention of the first referent, the comparative was manipulated such that it either would lead to anticipations for the subsequently named noun (Expected condition/True statement) or for the object that remained

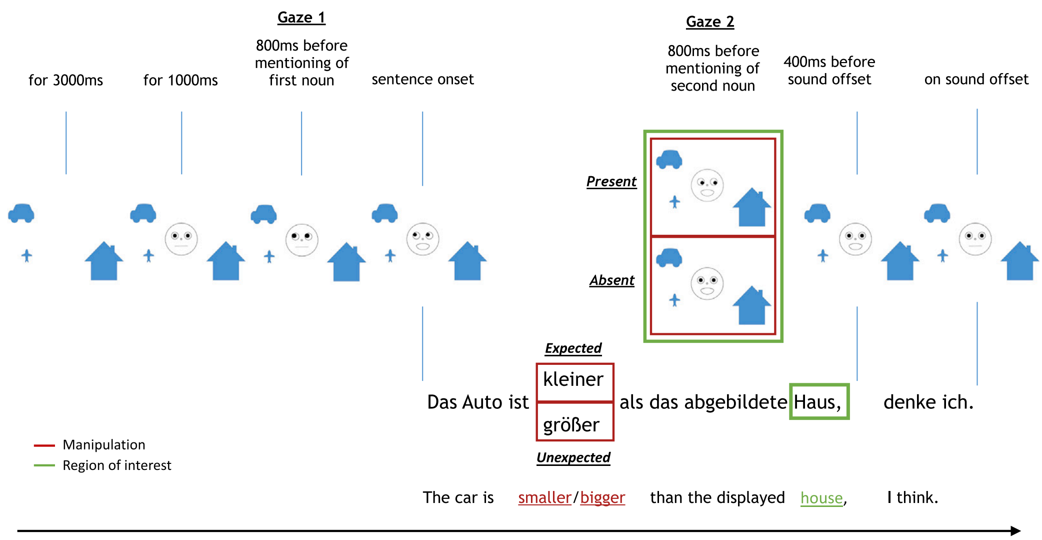


**Fig. 2.** Timeline of an experimental trial.

unmentioned in the sentence (Unexpected condition/False statement). Following the comparative, the second gaze cue occurred 800 ms prior to the mention of the second referent in the sentence (Gaze2). The gaze cue was manipulated in a way such that it was either directed toward the subsequently named object (Present condition) or redirected toward the center of the screen (Absent condition). 800 ms after the gaze action, the referent was mentioned. 400 ms prior to the end of the sentence, the gaze was redirected toward the center of the screen (for the Present gaze condition only) and the mouth closed on the sentence offset. Finally, participants were asked to indicate whether the sentence was true or false given the scene they saw by pushing the corresponding button.

Based on the four conditions, four lists were created following a Latin square design. Each list contained 152 experimental items (38 per condition) and 100 fillers. Fillers overall did not differ in sentence structure or gaze patterns, but were used to increase the number of referential gaze cues as well as their validity in order to reassure participants of their usefulness. Firstly, fillers were used to increase the proportion of true statements above chance to 65% (overall 164 true statements per list). This means that 88 of the fillers contained true statements and 12 false statements. Further, all fillers contained both referential gaze cues (toward NP1 and NP2). Thus, the overall distribution of sentence/gaze patterns was as follows: Combined with the 38 experimental items containing a present gaze cue and a true utterance, the 88 filler items increased the number of such trials (Present gaze — Expected continuation) to 50% (126). The number of false statements containing a gaze toward the target referent was increased to 20% (38 experimental items and 12 filler items). There were no fillers containing absent gaze, thus, both absent gaze conditions (true and false statements) constituted 15% (38) of the presented items per list each.

The stimuli were presented using E-prime (Version 2.0.10. Psychology Software Tools, Inc.). Each participant was seated in a sound-proof, electro-magnetically shielded chamber in front of a 24" Dell U2410 LCD monitor (resolution of 1280 × 1024 with a refresh rate of 75 Hz). The distance between the participant and the screen was always 114 cm in order to keep all objects in a 5° visual angle from the center of the screen. While the participants were prepared for the recording, they were presented with all objects that occurred throughout the experiment and their naming. The Alex voice of the CereVoice TTS was also used for the naming of the objects. After this, participants were presented with written instructions and completed six practice trials. The items were pseudo randomized for each list and presented in 5 blocks with breaks after each block. After each item, the participants were asked to indicate whether the sentence was true given the visual scene they were presented with by pressing one of two buttons. Answers were recorded using a Response Pad RB-834 (Cedrus Corporation). The experiment lasted approximately 60 min.

## 3. Participants

Thirty-four right-handed native speakers of German (Mean age: 24; Age range: [19, 33]; SD: 3.11; Female: 26) took part in the ERP experiment. Three participants had to be removed due to technical errors. Further 7 participants were removed due to the 30% threshold for the rejection rate per condition for participant exclusion due to eye-movements and other artifacts, similar to previous experiments. Overall, the remaining participants performed 5.2% eye movements on average. While there were significantly more eye movements in the noun region than in the gaze region ($\beta = 0.92, SE = 0.28, t = 3.26, p < .01, CI = [0.3433; 1.4901]$), there were no significant differences between experimental conditions for either region. Participants performed well in the task with an average of 97.5% of correct answers. There was no difference in accuracy between conditions ($F(3, 69) = 1.93, p = .15$). Overall, the criteria led to the removal of the data of 10 participants. After artifact rejection and participant exclusion 94.8% of the trials on average per participants were included in the analyses. Participants gave informed consent. All participants had normal or corrected-to-normal vision and had no hearing problems. All participants were compensated with €20 for their participation.

## 4. Data analysis and results

The EEG was recorded by 24 Ag/AgCl[1] scalp electrodes (actiCAP, BrainProducts) and amplified with a BrainAmp (BrainVision) amplifier. Electrodes were placed according to the 10–20 system (Sharbrough et al., 1991). Impedances were kept below 5k$\Omega$. The ground electrode was placed at AFz. The signal was referenced online to the reference electrode FCz and digitized at a sampling rate of 500 Hz. The EEG files were re-referenced offline to the average of the mastoid electrodes. The

[1] This excludes the electrodes used for the electrooculogram and offline re-reference: Fp1, Fp2, T7, T8, TP9, TP10, PO9 and PO10.

**Table 2**
Contrast matrix as embedded in the factor Condition in experiment 3.

| Condition | c1 | c2 | c3 |
|---|---|---|---|
| PE | −1/2 | −1/2 | 0 |
| PU | −1/2 | 1/2 | 0 |
| AE | 1/2 | 0 | −1/2 |
| AU | 1/2 | 0 | 1/2 |

P — Present gaze , A — Absent gaze , E — Expected continuation , U — Unexpected continuation
c1 — PvsA , c2 — P:EvsU , c3 — A:EvsU

**Table 3**
Contrast matrix as embedded in the factor Longitude.

| Longitude | l1 | l2 |
|---|---|---|
| Frontal | −2/3 | 0 |
| Central | 1/3 | −1/2 |
| Parietal | 1/3 | 1/2 |

l1 — FvsCP , l2 — CvsP

horizontal electrooculogram (EOG) was monitored with two electrodes placed at the right and left outer canthi of each eye and the vertical EOG with two electrodes below both eyes paired with Fp1 and Fp2. During recording an anti-aliasing low-pass filter of 250 Hz was used. The EEG data was band pass filtered offline at 0.01–40 Hz in order to attenuate skin potentials and other low voltage changes as well as line noise and EMG noise (Luck, 2014). Single-participant averages were computed for a 900 ms window per condition relative to the acoustical onset of the noun following the manipulated gaze cue and the manipulated gaze cue itself. All segments were aligned to a 100 ms pre-stimulus baseline. The data was semi-automatically screened offline for electrode drifts, amplifier blocking, eye-movements and muscle artifacts.

The recorded data was processed in Matlab (MATLAB, 2017), utilizing the EEGLAB plugin (Version 14.1.1b) (Delorme & Makeig, 2004) in combination with the ERPLAB extension (Version 6.1.4) (Lopez-Calderon & Luck, 2014).

The data (Jachmann et al., 2022) of the remaining 24 participants (Mean age: 24.3; Age range: [19, 31]; SD: 3.2; Female: 24) was exported from MATLAB in a format processable by R including item information for a time window between −100 to 800 ms relative to the onset of the two regions of interest (ROIs): The onset of the gaze cue toward the second noun, and the onset of said noun. Analyses were performed in R (Version 3.6.1) by fitting Linear Mixed-Effects Models using the lme4 (Bates et al., 2015b) package (Version 1.1–10). $\beta$-estimates, standard error and $t$-value are reported as well as the p value for significant effects and confidence intervals. The p values were extracted utilizing the lmerTest package (Version 3.0–1). The confidence intervals were extracted utilizing the profile function of the stats package (Version 3.6.1). Model fit was used to determine the reported models by model comparison utilizing the anova function also implemented in the stats package such that models with a lower AIC were preferred.

In a primary step, models with maximal random structure were fitted following (Barr et al., 2013). Based on the results from the previous experiments (Jachmann et al., 2019), the hypothesized utilization of gaze as an integral element of the communicative system, as well as evident from visual inspection of the data, a main effect of Gaze Presence was predicted. In order to investigate our hypothesis that gaze is indeed part of the communicative signal, and the corresponding prediction that the evaluation of expectations should shift from the noun region (when gaze is absent) to the preceding gaze cue (when gaze is present), we utilized a nested structure of the contrasts. This led to the following coding of the contrasts: All four conditions, namely Present gaze — Expected continuation (PE), Present gaze — Unexpected continuation (PU), Absent gaze — Expected continuation (AE) and Absent gaze — Unexpected continuation (AU), were encoded in a single factor. The embedded contrast structure was coded so that Expectancy was nested under Gaze Presence. Therefore, the factor included comparisons firstly for the Presence of gaze (c1 – Present versus Absent — PvsA) followed by a comparison of the Expectancy of the continuation under each of the two Gaze Presence factor levels (c2 – Present: Expected versus Unexpected — P:EvsU and c3 — Absent: Expected versus Unexpected — A:EvsU). The embedded contrast matrix can be seen in Table 2. Importantly, the nested structure most directly tests our hypotheses, which emphasize the determination of whether gaze is being treated as a fully informative part of the communicative signal. This would entail a shift in the region in which effects appear based on the presence of informative gaze (c2 for the gaze region and c3 for the noun region). It further addresses two issues that would be introduced by an interaction approach: Firstly, a classic interaction of the factors is not entirely meaningful in the gaze region as the comparison of Expected and Unexpected is not defined under Absent gaze. Thus, only the comparisons of gaze Presence and Absence (c1), as well of Expected and Unexpected under Present gaze (c2) were meaningful comparisons. Maintaining the same contrast structure for both regions further allows for a more straight froward comparison of the two regions. Secondly, as the noun region is immediately following the gaze region, the predicted effects in the gaze region when gaze was present in contrast to a lack of effects when gaze was absent would entail differences for the baseline of the noun region. Based on those differences, potential interactions could lead to false conclusions which are avoided by comparing only conditions that had the same baseline (c2 and c3). For comparison purposes and in the interest of completeness, Appendix B provides a corresponding interaction analysis.

Additionally, Longitude was added as a fixed effect in order to attest for scalp distribution of potential effects. For this factor the electrodes were grouped into 3 ROIs for frontal (F3, Fz, F4, FC5, FC1, FC2, FC6), central (C3, Cz, C4, CP5, CP1, CP2, CP6) and posterior (P7, P3, Pz, P4, P8, O1, O2) distributions. Contrasts were directly embedded into the factor with two comparisons. Firstly, frontal against centro-parietal electrodes (l1 — FvsCP) and, secondly, central against parietal electrodes (l2 — CvsP). Table 3 summarizes the corresponding contrast matrix. The full random structure including the interaction between Condition and Longitude under subject and item led to singular fit warnings in every time-window.

We followed the parsimonious mixed models approach as described by Bates et al. (2015a) utilizing PCAs to reduce the random structure of the models. The resulting random structure used in the final models is reported for each model.

### *4.1. Gaze cue preceding the target noun*

In order to isolate the involved components and to establish the time-windows for the analyses, the approach utilizing difference waves was followed as proposed by Kappenman and Luck (2016) and previously used by Jachmann et al. (2019) to identify the time-windows in the noun region. A Present-minus-Absent gaze difference wave was created as well as an Unexpected-minus-Expected difference wave for both levels of gaze presence that can be found in Fig. 3. Specifically, we identified two time-windows in the gaze region showing effects to the manipulations. The first time-window spanned from 250–800 ms for the comparison of Present with Absent gaze cues and the second time-window spanned from 250–450 ms for the comparison of Expected with Unexpected gaze cues when gaze was Present. In the following, the first time-window will be referred to as the P300 time-window (250-800 ms) rather than the P600 time-window based on its similarity in temporal distribution with the P300 component. The second time-window will be referred to as the N400 time-window (250–450 ms).

#### *4.1.1. P300 (250–800 ms)*

In the P300 time-window, there was a main effect of Gaze Presence ($\beta = -1.7, SE = 0.33, t = -5.1, p < .001, CI = [-2.3684; -0.9539]$) with a more positive deflection for Present gaze as compared to absent gaze as can be seen in Fig. 4. For a summary of the model results, see Table 4.

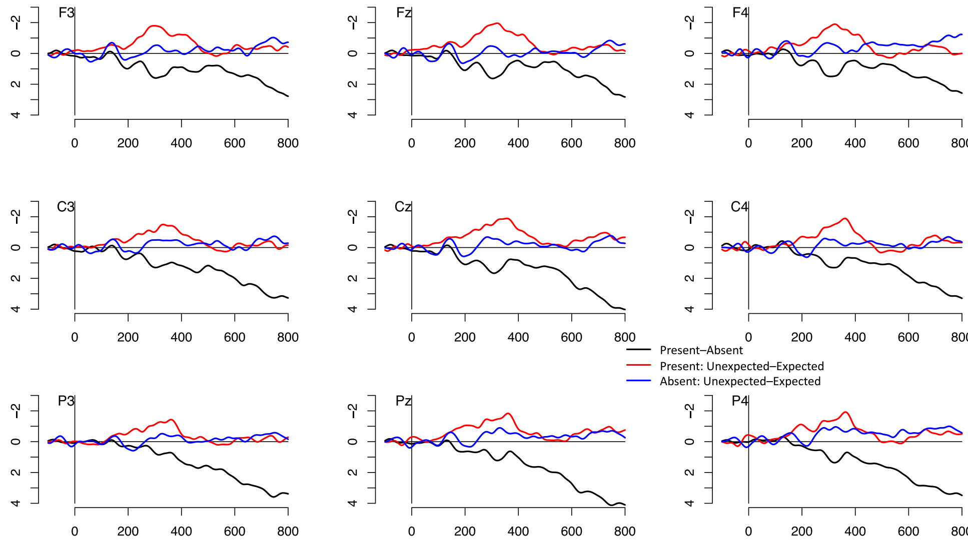


**Fig. 3.** Difference waves of Present-minus-Absent gaze cues (black), Unexpected-minus-Expected under Present gaze cues (red) and Unexpected-minus-Expected under Absent gaze cues (blue). The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20 Hz for presentation purposes only. (Negativity is plotted upward.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

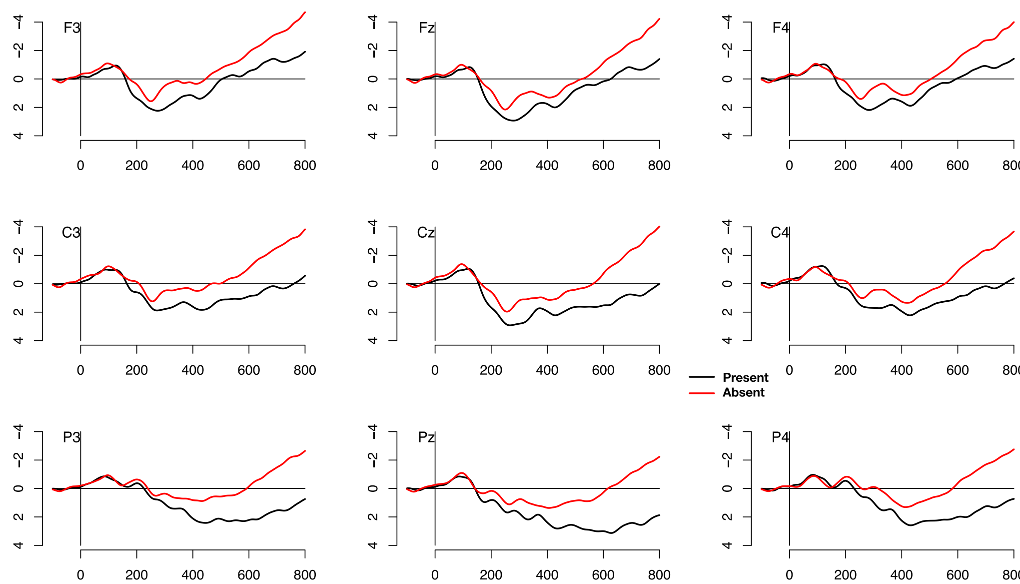


**Fig. 4.** ERP time-locked to the Gaze Cue Onset separated by Gaze Presence (Present: black, Absent: red). The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20 Hz for presentation purposes only. (Negativity is plotted upward.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

#### *4.1.2. N400 (250–450 ms)*

The N400 time-window again showed a main effect of Gaze Presence with a stronger positive deflection for Present gaze ($\beta = -1.05, SE = 0.31, t = -3.4, p < .01, CI = [-1.6385; -0.4691]$). Additional to this main effect, however, an effect of Expectancy was found when gaze was Present ($\beta = -1.06, SE = 0.43, t = -2.5, p < .05, CI = [-2.0653; -0.3115]$) with a more negative deflection for gaze cues toward the Unexpected referent as can be seen in Fig. 5. This effect was strongest in fronto-central electrodes ($\beta = 0.45, SE = 0.19, t = 2.5, p < .05, CI = [-0.0232; 0.8911]$). (For a summary of the model results, see Table 5.)

### *4.2. Noun region*

For the analyses of the noun following the gaze cue, the same time-windows were used that were established in Jachmann et al. (2019). More precisely, models were fitted for the PMN time-window (150–300 ms), the N400 time-window (300–450 ms), and the P600 time-window (600–800 ms). Fig. 6 shows the ERPs in the noun region when gaze was Present, whereas Fig. 7 shows the ERPs when gaze was Absent. The split was made due to the extended positivity on the gaze cue when gaze is Present, which overlaps with the baseline of the noun region. (For comparison, see Fig. 8, which displays the ERPs spanning over both ROIs.) Also evident from the ERPs is that the distinction between the PMN and N400 is not visibly detectable. This is however in accordance with the findings of Jachmann et al. (2019) where the distinction of the two components was only visibly — and statistically — detectable for the Neutral condition, where no anticipation violation was present. Rather, the Neutral condition contained no possibility for preliminary anticipation such that the acoustic input could be used to select the correct target, but was not violating any anticipation of the referential expression's realization. The negative deflection found in the current study, however, strongly resembles the negative deflection found for the Incongruent condition in Jachmann et al. (2019), which

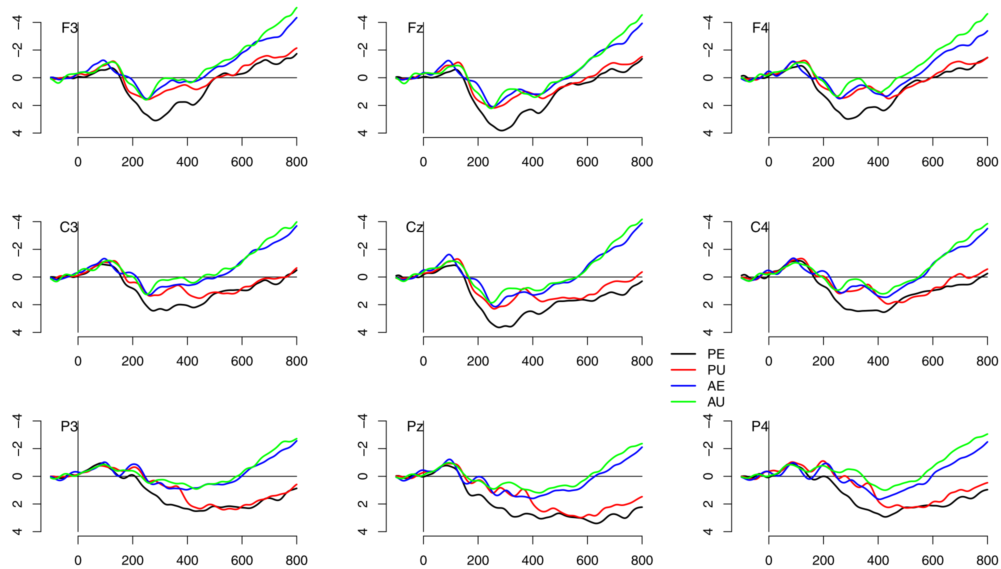


**Fig. 5.** ERP time-locked to the Gaze Cue Onset separated by the 4 experimental conditions (Present-Expected: black, Present-Unexpected: red, Absent-Expected: blue and Absent-Unexpected: green). The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20 Hz for presentation purposes only. (Negativity is plotted upward.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

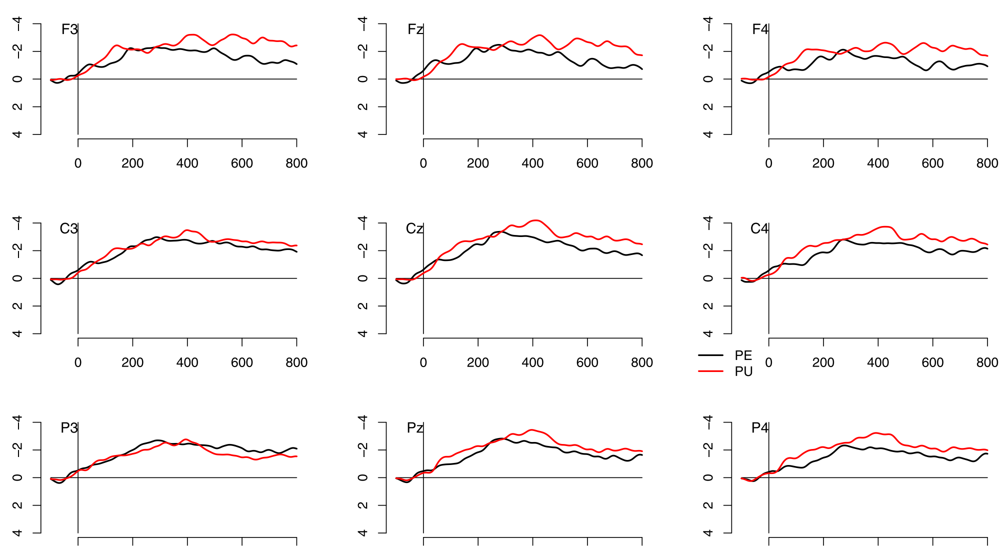


**Fig. 6.** ERP time-locked to the Noun Onset separated by Expectancy under Present Gaze (Present-Expected: black, Present-Unexpected: red.) The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20 Hz for presentation purposes only. (Negativity is plotted upward.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

— much as the Unexpected condition in the current study — contained an expectation violation. That violation was interpreted to lead to a stronger N400 effect that, due to component overlap, leads to a non-distinguishable negative deflection. Notably, the time-windows for the negative deflections in the current study (PMN: 150–300 ms and N400: 300–450 ms) is the same as in the previous studies.

*4.2.1. PMN (150–300 ms)*

In the PMN time-window, there was a main effect of Gaze Presence ($\beta = 0.68, SE = 0.33, t = 2.0, p < .05, CI = [-0.0085; 1.4437]$) with a more positive deflection when a gaze cue was Present. This effect can most likely be attributed to the preceding gaze region, which showed a more positive movement of the ERPs in the P300 time-window, which overlaps with the baseline for the noun region. Above and beyond this effect however, the comparison of Expected and Unexpected continuations, which differ in the expectancy of the gaze and consequent noun, led to a significant difference only when the preceding gaze toward the object was absent (conditions AE and AU). In this time window, unexpected nouns led to a more negative deflection when gaze was absent ($\beta = -1.36, SE = 0.50, t = -2.7, p < .01, CI = [-2.4024; -0.5467]$) as can be seen in Fig. 7. As reported above, no such difference was found in the noun region when preceding gaze was present as can be seen in Fig. 6. (For a summary of the model results, see Table 6.)

*4.2.2. N400 (300–450 ms)*

In the N400 time-window, there again was an effect of Gaze Presence ($\beta = 2.18, SE = 0.54, t = 4.1, p < .001, CI = [1.0192; 3.4835]$) similar to the preceding time-window. This effect was only present in centro-parietal regions ($\beta = 0.77, SE = 0.34, t = 2.2, p < .05, CI = [0.0852; 1.4853]$), supporting a distinction from the globally distributed preceding PMN component. The comparison of Expected and Unexpected continuations when gaze was absent shows a negative deflection

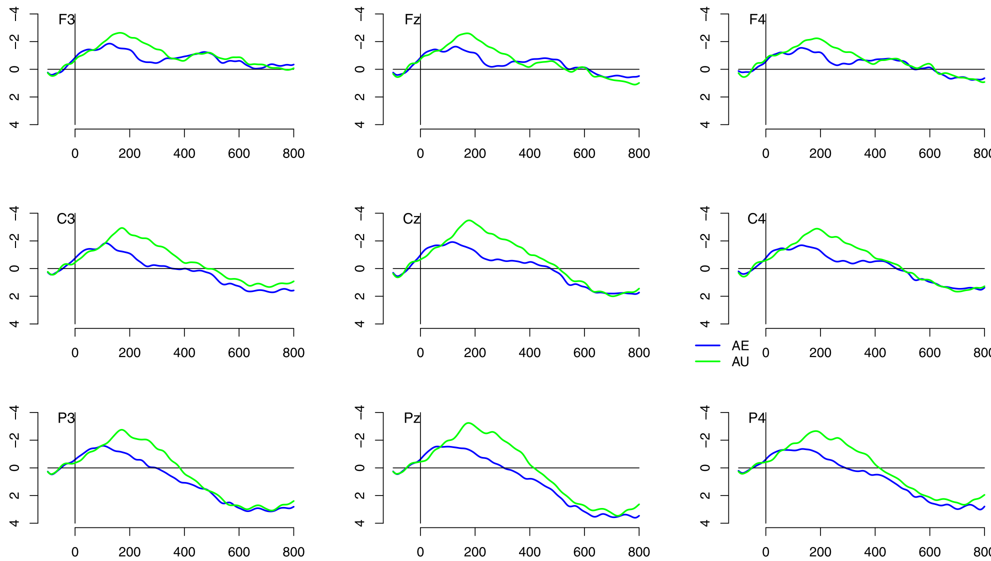


**Fig. 7.** ERP time-locked to the Noun Onset separated by Expectancy under Absent Gaze. (Absent-Expected: blue and Absent-Unexpected: green.) The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20 Hz for presentation purposes only. (Negativity is plotted upward.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

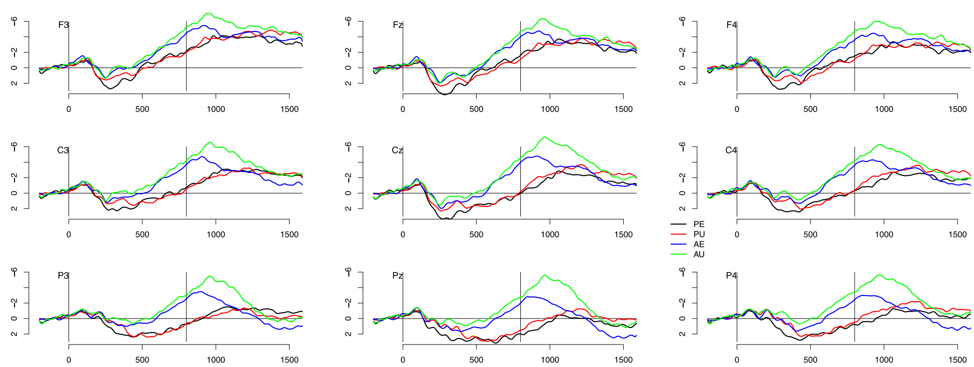


**Fig. 8.** ERP time-locked to the Gaze Cue Onset separated by the 4 experimental conditions, spanning both regions of interest. (Present-Expected (black), Present-Unexpected (red), Absent-Expected (blue) and Absent-Unexpected (green). The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20 Hz for presentation purposes only. (Negativity is plotted upward.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 4**
Fixed effects for P300 time-window in the Gaze region.

| Fixed effect | $\beta$-estimate | Std. error | t value | sig. |
|---|---|---|---|---|
| (Intercept) | 0.22017 | 0.32183 | 0.68 | |
| c1 | −1.70320 | 0.33296 | −5.11 | *** |
| c2 | −0.56046 | 0.40668 | −1.38 | |
| c3 | −0.35322 | 0.36368 | −0.97 | |
| l1 | 0.54351 | 0.18282 | 2.97 | ** |
| l2 | −0.13531 | 0.15963 | −0.85 | |
| c1:l1 | −0.33218 | 0.24392 | −1.36 | |
| c1:l2 | 0.09708 | 0.16250 | 0.60 | |
| c2:l1 | 0.18504 | 0.36154 | 0.51 | |
| c2:l2 | 0.29165 | 0.19525 | 1.49 | |
| c3:l1 | −0.03037 | 0.29115 | −0.10 | |
| c3:l2 | −0.05974 | 0.19475 | −0.31 | |

(. — $p < 0.1$ , * — $p < 0.05$ , ** — $p < 0.01$, *** — $p < 0.001$)
(c1 — PvsA , c2 — P:EvsU , c3 — A:EvsU , l1 — FvsCP , l2 — CvsP)

**Table 5**
Fixed effects for N400 time-window in the Gaze region in experiment 3.

| Fixed effect | $\beta$-estimate | Std. error | t value | sig. |
|---|---|---|---|---|
| (Intercept) | 0.97857 | 0.34574 | 2.83 | ** |
| c1 | −1.04827 | 0.30800 | −3.40 | ** |
| c2 | −1.06369 | 0.42890 | −2.48 | * |
| c3 | −0.37541 | 0.36370 | −1.03 | |
| l1 | −0.42562 | 0.19377 | −2.20 | * |
| l2 | −0.84948 | 0.19654 | −4.32 | *** |
| c1:l1 | 0.05419 | 0.21796 | 0.25 | |
| c1:l2 | 0.08392 | 0.16112 | 0.52 | |
| c2:l1 | 0.38143 | 0.30462 | 1.25 | |
| c2:l2 | 0.44953 | 0.19302 | 2.52 | * |
| c3:l1 | −0.14784 | 0.28182 | −0.52 | |
| c3:l2 | −0.03479 | 0.19253 | −0.18 | |

( . — $p < .1$ , * — $p < .05$ , ** — $p < .01$, *** — $p < .001$)
(c1 — PvsA , c2 — P:EvsU , c3 — A:EvsU , l1 — FvsCP , l2 — CvsP)

**Table 6**
Fixed effects for PMN time-window in the Noun region.

| Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|
| (Intercept) | −1.703833 | 0.275483 | −6.18 | *** |
| c1 | 0.676556 | 0.332749 | 2.03 | * |
| c2 | −0.049757 | 0.437639 | −0.11 | |
| c3 | −1.357151 | 0.495168 | −2.74 | ** |
| l1 | 0.016384 | 0.185817 | 0.09 | |
| l2 | 0.557507 | 0.111996 | 4.98 | *** |
| c1:l1 | 0.000697 | 0.251408 | 0.00 | |
| c1:l2 | −0.164593 | 0.153343 | −1.07 | |
| c2:l1 | −0.014542 | 0.311906 | −0.05 | |
| c2:l2 | 0.084014 | 0.188377 | 0.45 | |
| c3:l1 | −0.383718 | 0.303006 | −1.27 | |
| c3:l2 | 0.315112 | 0.218423 | 1.44 | |

( . — $p < 0.1$ , ∗ — $p < 0.05$ , ∗∗ — $p < 0.01$, ∗∗∗ — $p < 0.001$)
(c1 — PvsA , c2 — P:EvsU , c3 — A:EvsU , l1 — FvsCP , l2 — CvsP)

**Table 7**
Fixed effects for N400 time-window in the Noun region.

| Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|
| (Intercept) | −1.35214 | 0.37409 | −3.61 | *** |
| c1 | 2.18637 | 0.53518 | 4.08 | *** |
| c2 | −0.52315 | 0.47827 | −1.09 | |
| c3 | −0.59339 | 0.54488 | −1.09 | |
| l1 | 0.20288 | 0.22387 | 0.91 | |
| l2 | 0.71760 | 0.12416 | 5.78 | *** |
| c1:l1 | 0.76503 | 0.34281 | 2.23 | * |
| c1:l2 | −0.05357 | 0.26494 | −0.20 | |
| c2:l1 | 0.13393 | 0.36788 | 0.36 | |
| c2:l2 | 0.33044 | 0.21316 | 1.55 | |
| c3:l1 | −0.69905 | 0.32493 | −2.15 | * |
| c3:l2 | −0.00844 | 0.24454 | −0.03 | |

( . — $p < .1$ , ∗ — $p < .05$ , ∗∗ — $p < .01$, ∗∗∗ — $p < .001$)
(c1 — PvsA , c2 — P:EvsU , c3 — A:EvsU , l1 — FvsCP , l2 — CvsP)

for Unexpected continuations only in centro-parietal regions ($\beta = -0.70, SE = 0.34, t = -2.2, p < .05, CI = [-1.3555; 0.0221]$). (For comparison, see Fig. 7.) A summary of the model results can be found Table 7.

*4.2.3. Distinction of PMN and N400*

In order to further support the distinction of the two components in the noun region, beyond the difference in distribution, we analyzed the PMN time-window in the gaze region: Similar to the reported N400 time-window for the gaze cue, the PMN window as well shows a stronger positive deflection for Present gaze ($\beta = -0.59, SE = 0.28, t = -2.1, p < .05, CI = [-1.1652; -0.0729]$). However, importantly, there was no significant difference in this time-window for Expectancy, independent from whether gaze was Present or Absent ($\beta = -0.75, SE = 0.4, t = -1.87, p = .06$ and $\beta = -0.08, SE = 0.33, t = -0.24, p = .8$ respectively), supporting a difference in the time-course of effects between the two regions. A summary of the model results can be found Table 8.

*4.2.4. P600 (600–800 ms)*

As for the previous time-windows, the P600 time-window also shows a main effect of Gaze Presence ($\beta = 3.43, SE = 0.57, t = 6.0, p < .001, CI = [1.9730; 4.7429]$) which is more prominent in centro-parietal regions ($\beta = 1.63, SE = 0.48, t = 3.4, p < .01, CI = [0.8295; 2.5942]$). Different from Jachmann et al. (2019), in this time-window, no effect of Expectancy was found. (For a summary of the model results, see Table 9.) Importantly, the positivity in response to the Gaze Presence factor (c1) occurs for Present gaze in the gaze region (P300 time-window) and for Absent gaze in the noun region as evident from the opposite polarities of the effects for the comparison between Present and Absent gaze (c1). This swap of positivity-inducing condition strongly indicates the validity of this result. However, as we argued before that this comparison might not be necessarily informative based on the baseline differences due to the preceding positivity in responses to the gaze cue, we ran an additional analysis of this time-window. For this analysis, we used the data spanning both the gaze cue and noun region, with a baseline correction only done relative to the gaze cue onset. In other words, we extracted a time-window spanning 1600 ms from the gaze cue onset with an additional 100 ms before the gaze cue for baseline correction. Hence, the time-window for the P600 in the noun region was 1400–1600 ms relative to the gaze cue onset. The Linear Mixed-Effects Model fitted to this time-window supported the above finding, such that an effect of Gaze Presence was found in central-parietal regions ($\beta = 0.98, SE = 0.42, t = 2.4, p < .05, CI = [0.1625; 1.8263]$). Again, as in the analysis based on the data that was baseline-corrected on the noun, no effect of Expectancy was found. (For a summary of the model results, see Table 10.)

**Table 8**
Fixed effects for 150–300 ms time-window in the Gaze region in experiment 3.

| Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|
| (Intercept) | 0.24255 | 0.34574 | 1.19 | |
| c1 | −0.58471 | 0.27697 | −2.11 | * |
| c2 | −0.74988 | 0.40154 | −1.87 | |
| c3 | −0.08071 | 0.33307 | −0.24 | |
| l1 | −1.22910 | 0.21512 | −5.71 | *** |
| l2 | −1.26837 | 0.16012 | −7.92 | *** |
| c1:l1 | 0.32916 | 0.19457 | 1.69 | |
| c1:l2 | 0.27112 | 0.12485 | 2.17 | * |
| c2:l1 | 0.18793 | 0.30935 | 0.61 | |
| c2:l2 | 0.19708 | 0.17694 | 1.11 | |
| c3:l1 | −0.07108 | 0.30547 | −0.23 | |
| c3:l2 | −0.15134 | 0.17649 | −0.86 | |

( . — $p < .1$ , ∗ — $p < .05$ , ∗∗ — $p < .01$, ∗∗∗ — $p < .001$)
(c1 — PvsA , c2 — P:EvsU , c3 — A:EvsU , l1 — FvsCP , l2 — CvsP)

## 5. Discussion

The present study set out to establish whether gaze could be utilized not only to help listeners form expectations about upcoming referring expressions, as shown in previous experiments (Jachmann et al., 2019), but also convey referential information that can be used to either confirm or disconfirm expectations induced by the prior linguistic context. In their study, gaze was used to establish expectations for a subsequent noun. Here, the gaze cue and the subsequent noun both can potentially be utilized to confirm or falsify expectations based on the preceding linguistic context. While the noun is always present to fill this role, a preceding gaze can either be present or absent. As a consequence, contextually established expectations for a specific referent could be confirmed (or falsified) by either the gaze cue (if present) or the subsequent noun.

In the following, the findings for the ERP components will be summarized and interpreted according to the underlying mechanisms they are indexing. An overview of the reported effects can be found in Table 11.

*5.1. PMN*

In line with the interpretation of the PMN effect as an informativity driven phonological matching mechanism (Jachmann et al., 2019), visual information as provided by the gaze cue does not show this effect

**Table 9**
Fixed effects for P600 time-window in the Noun region.

| Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|
| (Intercept) | −0.16471 | 0.38735 | −0.42 | |
| c1 | 3.43266 | 0.57132 | 6.01 | *** |
| c2 | −0.48065 | 0.54607 | −0.88 | |
| c3 | −0.20360 | 0.51067 | −0.40 | |
| l1 | 0.76357 | 0.26404 | 2.89 | ** |
| l2 | 0.55883 | 0.23440 | 2.38 | * |
| c1:l1 | 1.62503 | 0.48165 | 3.37 | ** |
| c1:l2 | −0.02816 | 0.27660 | −0.10 | |
| c2:l1 | 0.75172 | 0.39708 | 1.89 | . |
| c2:l2 | 0.47607 | 0.24724 | 1.93 | . |
| c3:l1 | −0.31782 | 0.35783 | −0.89 | |
| c3:l2 | −0.06857 | 0.30284 | −0.23 | |

( . — $p < 0.1$ , ∗ — $p < 0.05$ , ∗∗ — $p < 0.01$, ∗∗∗ — $p < 0.001$)
(c1 — PvsA , c2 — P:EvsU , c3 — A:EvsU , l1 — FvsCP , l2 — CvsP)

**Table 10**
Fixed effects for P600 time-window in the Noun region (baseline corrected on the preceding gaze cue).

| Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|
| (Intercept) | −1.2701 | 0.5728 | −2.22 | * |
| c1 | 0.6092 | 0.4907 | 1.24 | |
| c2 | −0.5395 | 0.7017 | −0.77 | |
| c3 | −0.9495 | 0.6338 | −1.50 | |
| l1 | 2.0814 | 0.3892 | 5.35 | *** |
| l2 | 1.0639 | 0.3747 | 2.84 | ** |
| c1:l1 | 0.9810 | 0.4168 | 2.35 | * |
| c1:l2 | 0.2601 | 0.2575 | 1.01 | |
| c2:l1 | 0.6520 | 0.5148 | 1.27 | |
| c2:l2 | 0.5645 | 0.3288 | 1.72 | . |
| c3:l1 | −0.3563 | 0.3354 | −1.06 | |
| c3:l2 | −0.1752 | 0.3252 | −0.54 | |

(. — $p < 0.1$ , ∗ — $p < 0.05$ , ∗∗ — $p < 0.01$, ∗∗∗ — $p < 0.001$)
(c1 — PvsA , c2 — P:EvsU , c3 — A:EvsU , l1 — FvsCP , l2 — CvsP)

as supported by the difference wave approach. This suggests that the PMN is only elicited in response to the phonological information in the acoustic signal, and is not sensitive to any retrieved phonological form of the gazed-at-referent. But, in the noun region, the PMN modulation was observed, similar to previous findings. Thus, this modulation of the PMN expresses the processing of an acoustically perceived phoneme that was unexpected given the context. Crucially, when gaze could be used to disconfirm the previously formed expectations, no PMN effect was observed. Thus, an informative gaze cue could lead to a shift in expectations upon hearing the noun corresponding to the gazed-at-object. In other words, the gaze cue could be utilized to disconfirm the expectations formed based on the preceding linguistic content, resulting in the retrieval of the gazed-at object as well as the revision of the mental representation of the situation. Additionally, as it was argued that expectations about the upcoming referent are derived from the current state of the mental representation of the situation (Jachmann et al., 2019), the adjustment of that representation should also entail adjustments to the expectations about the upcoming referent. This would support the view of a more fluid system of information integration that is constantly monitored and updated based on linguistic and situational information.

### 5.2. N400

In both the gaze region as well as the noun region, a significant negativity was found for unexpected gaze and nouns respectively. Importantly, the N400 in the noun region was only present when there was no preceding gaze cue. Even though the time-window for the N400 in the gaze region differs slightly from the time-window in the noun region (250–450 ms and 300–450 ms respectively), it is similar to other reported time-windows of N400s in response to pictures (e.g., 270–420ms (Nigam et al., 1992)). Again in line with previous experiments, the N400 effect in the Unexpected conditions can be interpreted as expectation violations resulting in more effortful retrieval of the referent. As with the PMN time-window, the N400 effect in the noun region is only present when a preceding gaze cue was absent. Similar to the explanation in the PMN time-window, this can be explained by a shift of expectations for the noun due to the preceding disconfirmation from the gaze cue. Hence, following the previous interpretation, the gaze cue was utilized to update the mental representation of the situation, such that the expectations derived from the current state of that representation shifted to the gazed-at object rather than the referent supported by the preceding linguistic content.

### 5.3. Positive deflections

The early positive deflection found in the gaze region that only occurs when gaze is present can be interpreted as the integration of the perceived information with the mental representation (Donchin, 1981). It can reasonably be interpreted as a P300 and, more precisely, a P3b (Donchin, 1981; Polich, 2007). This component has been reported to be elicited by task-related decision making (Verleger, 1997) and as expressing 'context model updating' (Donchin, 1981). This is in line with the assumption of a mental representation of the situation (e.g., Zwaan & Radvansky, 1998). Although, the P300 time-windows in the literature often differ based on the modalities of the respective experiments, these time-windows are broadly similar to the here reported time-window of 250–800 ms reported here.

In previous experiments (Jachmann et al., 2019), a positive deflection could be seen to the noun when the mental representation of the situation needed to be updated. This was only the case in the Incongruent conditions in those experiments. Unlike in the previous experiments however, here the positive deflection occurred in both conditions — Expected as well as Unexpected gaze. Similar to this finding in the gaze region, a positive deflection in the noun region is only present when preceding gaze was Absent. Again, this effect can be attributed to the integration in the mental representation of the situation. As both positive deflections — P300 (250–800 ms) on the gaze cue and P600 (600–800 ms) on the noun — are interpreted to index similar mechanisms of meaning integration/update (Bornkessel & Schlesewsky, 2006; Brouwer et al., 2012; Burkhardt, 2006), these findings are in line with the P600-as-P3 hypothesis (Bornkessel-Schlesewsky et al., 2011; Coulson et al., 1998; Sassenhagen et al., 2014). This interpretation is also in line with (Kuperberg et al., 2020), who suggested that late posterior positivities indicate conflicts of the input with the comprehender's communicative model. While our analyses do not reveal a significant interaction of experimental condition by topographic distribution in the gaze region — suggesting an overall global distribution — visual inspection does suggest a more parietal distribution of the positivity (which is also supported by the significant main effect of the l1 contrast — frontal against centro-parietal electrodes — indicating an overall more centro-parietal distribution). Further, in the noun region, we found an interaction of the l1 contrast with the c1 contrast — present against absent gaze — indicating a more centro-parietal distribution of the positivity when preceding gaze was absent. However, similar to the corresponding effect in the gaze region, the P600 in the noun region occurred in both Expectancy conditions under Absent gaze. While the effects in the earlier time-windows (PMN and N400) are in line with the predictions for gaze to be utilized to evaluate expectations based on the preceding linguistic content of an utterance, the findings in the positive deflections are not explicable by the hypothesized updating of a mental representation alone (Jachmann et al., 2019).

To possibly explain this difference, it is important to point out another difference in the role of the respective regions between the current experiment and the experiments in Jachmann et al. (2019).

**Table 11**
Summary of the statistics for both Regions Of Interest.

| Comparison | Gaze Region | | | Noun Region | | |
|---|---|---|---|---|---|---|
| | PMN | N400 | P300 | PMN | N400 | P600 |
| Present–Absent | – | – | + | – | – | – |
| Absent–Present | – | – | – | – | – | + |
| Present: Unexpected–Expected | – | + | – | – | – | – |
| Absent: Unexpected–Expected | – | – | – | + | + | – |

– — no effect , + — effect found

As previously stated, the main change was the different roles gaze and linguistic context took in the experiments. While in the previous experiments gaze could be utilized to form expectations about the linguistic content of the unfolding utterance, in the current experiment, gaze could instead be utilized to confirm expectations formed on the basis of the linguistic context of the utterance so far. In order to change these roles, the sentence structure was altered. While the sentences in the previous experiments were of the form *"Verglichen mit dem Auto, ist das Haus verhältnismäßig groß, denke ich" (Compared to the car, the house is relatively big, I think)*, the sentences in the current experiment were of the form *"Das Auto ist kleiner als das abgebildete Haus, denke ich" (The car is smaller than the displayed house, I think)*. This could be more generally expressed as a sequence of **NP1–Gaze–NP2–Comparative** for the former sentence and **NP1–Comparative–Gaze–NP2** for the latter sentence. When inspecting the order of the utterances' elements, it becomes clear that, beyond the shift in the role of the linguistic context and gaze, the point of sentence evaluation also shifts.

In the previous experiments, whether the sentence was true or false was only detectable once the comparative was uttered. Hence, neither the gaze region nor the noun region were affected by this evaluation process. In the current experiment, however, the new word order also led to an overlap of the expectation confirmation and the sentence evaluation. On the example of the sentence *'The car is smaller than the displayed house, I think'*, in order for the participants to decide whether the sentence was true or false given the visual context, they needed to know which of the objects was actually compared to the *'car'*. Gaze, when present, was used to confirm the expectations, however, it could additionally be used to evaluate the sentence at that point. The very same holds for the noun region in this experiment when gaze was absent.

### 5.4. Situational integration account

The current study provides evidence that gaze is not only utilized to form expectations about the upcoming linguistic signal but further conveys referential information that can be used to evaluate linguistic expectations. This suggests that speaker gaze serves as an integral part of the communicative signal, that can both elicit expectations and confirm them, in much the same manner as the spoken linguistic context, and referring expressions, respectively.

These findings are directly predicted by the previously proposed Situational Integration Account (Jachmann et al., 2019). The SIA builds upon the tradition of constraint-based theories regarding the utilization of linguistic, situational and world knowledge (e.g., Degen & Tanenhaus, 2019) to inform the incremental construction of the mental representation of an utterance. Importantly, however, the SIA further accommodates the dynamic, non-linguistic — but nonetheless communicative — cues provided by the speaker, with a mechanistic explanation of how they contribute to both incremental comprehension and expectations (Jachmann, 2020; Jachmann et al., 2019). Specifically, the SIA posits that such cues are not simply part of the situational context, in the way that objects (Tanenhaus et al., 1995) and events (Knoeferle et al., 2005) are typically treated, but rather that they are in fact part of the communicative signal. The SIA therefore directly predicts the present findings — namely that a gaze cue can not only elicit a similar neurophysiological response to that of an unspoken referent, but further that the presence of such a gaze cue can eliminate the response that would be typically evoked by the referent had gaze not been present. To our knowledge, these findings were not directly predicted by extant constraint-based accounts, prior to the SIA.

Further, such cues may not be limited to gaze alone but could also hold for other communicative elements of speaker behavior such as gestures, or possibly even speaker-external cues such as arrows as long as they are interpretable as relating to the communicative interaction (e.g., Staudte et al., 2014). However, we also argue that gaze still holds a distinguished role amongst those cues due to its constant availability in face-to-face interactions and the — at least in part — automatic, reflexive utilization of gaze by speakers in referential settings. That arrow cues may elicit similar results as the presented gaze cues is not in conflict with the SIA, particularly in settings where such iconic cues are context-situationally relevant and, crucially, reliably aligned with the linguistic signal. As such artificial cues are unlikely to occur in natural communicative interactions, it would presumably be necessary to first identify the source and reliability of such a cue in order for them to be treated as part of the communicative signal. In contrast, we have a life-long experience of not only observing speaker gaze as listeners, but also performing such gaze shifts as speakers.

Overall, the presented results are in support of and further extend the Situational Integration Account: In situated verbal interactions, any relevant information, such as the linguistic signal, the visual context, and speaker behavior, that can potentially facilitate comprehension is utilized incrementally as soon as it is available. That information is integrated into a mental representation of the situation, allowing for anticipation of upcoming elements of the unfolding utterance. These expectations are then confirmed or not by the subsequent signal — whether the utterance or gaze — leading to adjustments of the representation where and when necessary. The SIA assumes different mechanisms that are involved in the maintenance of the mental representation: A mechanism of matching the anticipated input with the perceived input that is inhibited by unexpected word forms (PMN) and that potentially alleviates semantic retrieval (as evident from shorter N400 effects following a PMN). A mechanism of semantic retrieval that is inhibited when expectations on the basis of the current state of the mental representation are not met, which is for example the case when an unexpected referent is mentioned or gazed at, but potentially also pointed to or otherwise indicated as part of the unfolding utterance (N400). Finally, a monitoring mechanism is assumed to enable a rapid reevaluation and adjustment of the mental representation. This monitoring system is also involved in the evaluation of the mental representation once enough information is provided to mediate decision making also in relation to the current situation (such as task, response-readiness, the formulation of a verbal response, etc.).

## CRediT authorship contribution statement

**Torsten Kai Jachmann:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Heiner Drenhaus:** Conceptualization, Methodology, Validation, Writing – review & editing. **Maria Staudte:** Conceptualization, Methodology, Validation, Writing – review & editing. **Matthew W. Crocker:** Conceptualization, Methodology, Validation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data is cited in the paper (OSF: https://osf.io/uzktp/).

## Acknowledgments

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) — Projektnummer 232722074 — SFB 1102/Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — Project-ID 232722074 — SFB/CRC 1102.

## Appendix A. Model code as used in R

The following code provides the full coding of the reported models for each time-window as utilized using the lme4 package in R.

(c1 — PvsA, c2 — P:EvsU, c3 — A:EvsU, l1 — FvsCP, l2 — CvsP)

```
final.Gaze.P3.model = lmer(eeg ~ 1 + (c1 + c2 + c3)*(l1 + l2) +
            (1 + c1 + c2 + c3 + l1 + l2 | subject) +
            (0 + c1:l1 | subject) +
            (0 + c1:l2 | subject) +
            (0 + c2:l1 | subject) +
            (0 + c3:l1 | subject) +
            (1 + c1 + c2 + c3 + l1 | item) +
            (0 + l2 | item) +
            (0 + c1:l1 | item) +
            (0 + c2:l1 | item) +
            (0 + c3:l1 | item),
          REML=FALSE, data=Gaze.P3.Data,
          control = lmerControl(calc.derivs=FALSE),
          na.action = na.omit)

final.Gaze.N400.model = lmer(eeg ~ 1 + (c1 + c2 + c3)*(l1 + l2) +
            (1 + c1 + c2 + c3 + l1 + l2 | subject) +
            (0 + c1:l1 | subject) +
            (0 + c1:l2 | subject) +
            (0 + c2:l1 | subject) +
            (1 + c1 + c2 + c3 + l1 | item) +
            (0 + l2 | item) +
            (0 + c1:l1 | item) +
            (0 + c2:l1 | item) +
            (0 + c3:l1 | item),
          REML=FALSE, data=Gaze.N400.Data,
          control = lmerControl(calc.derivs=FALSE),
          na.action = na.omit)

final.Noun.PMN.model = lmer(eeg ~ 1 + (c1 + c2 + c3)*(l1 + l2) +
            (1 + c1 + c2 + c3 + l1 | subject) +
            (0 + l2 | subject) +
            (0 + c1:l1 | subject) +
            (0 + c1:l2 | subject) +
            (0 + c2:l1 | subject) +
            (0 + c3:l1 | subject) +
            (0 + c3:l2 | subject) +
            (1 + c1 + c2 + c3 + l1 | item) +
            (0 + l2 | item) +
            (0 + c1:l1 | item) +
            (0 + c2:l1 | item) +
            (0 + c3:l1 | item),
          REML=FALSE, data=Noun.PMN.Data,
          control = lmerControl(calc.derivs=FALSE),
          na.action = na.omit)

final.Gaze.PMN.model = lmer(eeg ~ 1 + (c1 + c2 + c3)*(l1 + l2) +
            (1 + c1 + c2 + c3 + l1 + l2 | subject) +
            (0 + c1:l1 | subject) +
            (0 + c2:l1 | subject) +
            (0 + c3:l1 | subject) +
            (1 + c1 + c2 + c3 + l1 + c1:l1 + c2:l1 + c3:l1 | item) +
            (0 + l2 | item),
          REML=FALSE, data=Noun.PMN.Data,
          control = lmerControl(calc.derivs=FALSE),
          na.action = na.omit)

final.Noun.N400.model = lmer(eeg ~ 1 + (c1 + c2 + c3)*(l1 + l2) +
            (1 + c1 + c2 + c3 + l1 + l2 | subject) +
            (0 + c1:l1 | subject) +
            (0 + c1:l2 | subject) +
            (0 + c2:l1 | subject) +
            (0 + c2:l2 | subject) +
            (0 + c3:l1 | subject) +
            (0 + c3:l2 | subject) +
            (1 + c1 + c2 + c3 + l1 + c1:l1 + c2:l1 + c3:l1 | item),
          REML=FALSE, data=Noun.N400.Data,
          control = lmerControl(calc.derivs=FALSE),
          na.action = na.omit)

final.Noun.P600.model = lmer(eeg ~ 1 + (c1 + c2 + c3)*(l1 + l2) +
            (1 + c1 + c2 + c3 + l2 + c1:l2 + c3:l2 | subject) +
            (0 + l1 | subject) +
            (0 + c1:l1 | subject) +
            (0 + c2:l1 | subject) +
            (0 + c3:l1 | subject) +
            (1 + c1 + c2 + c3 + l1 + c1:l1 + c2:l1 + c3:l1 | item),
          REML=FALSE, data=EXP3.Noun.P600.Data,
          control = lmerControl(calc.derivs=FALSE),
          na.action = na.omit)
```

## Appendix B. Interaction models

For completeness purposes and for comparison with the analyses utilizing nested contrast structures (see Section 4), this Appendix provides interaction models and the respective follow-up splits for the analyses of the noun region. However, effects here should be taken with caution as these do not account for any potential confound introduced by the preceding gaze cue. Recall that the noun immediately follows the gaze cue and, thus, the baseline correction conflates the baseline differences introduced by the present/absent gaze condition split (see Tables B.1–B.7).

**Table B.1**
Fixed effects for PMN time-window in the Noun region — Interaction Model.

| Comp. | Fixed effect | $\beta$-estimate | Std. error | t value | sig. |
|---|---|---|---|---|---|
| (Intercept) | | −1.703310 | 0.275735 | −6.18 | *** |
| c1 | g1 | 0.675029 | 0.333656 | 2.02 | * |
| | e1 | −0.703295 | 0.335196 | −2.10 | * |
| l1 | l1 | 0.017162 | 0.185856 | 0.09 | |
| l2 | l2 | 0.558038 | 0.112338 | 4.97 | *** |
| | g1:e1 | −1.304720 | 0.651265 | −2.00 | * |
| c1:l1 | g1:l1 | 0.001750 | 0.252418 | 0.01 | |
| c1:l2 | g1:l2 | −0.163372 | 0.153153 | −1.07 | |
| | e1:l1 | −0.202348 | 0.208780 | −0.97 | |
| | e1:l2 | 0.201074 | 0.132340 | 1.52 | |
| | g1:e1:l1 | −0.372259 | 0.449539 | −0.83 | |
| | g1:e1:l2 | 0.233338 | 0.264680 | 0.88 | |

The column Comp. provides information on which effect in the nested structure the reported effect can be compared to.
( . — $p < 0.1$ , * — $p < 0.05$ , ** — $p < 0.01$, *** — $p < 0.001$)
(g1 — Present vs Absent Gaze , e1 — Expected vs Unexpected , l1 — FvsCP , l2 — CvsP)

**Table B.2**
Fixed effects for PMN time-window in the Noun region — Present Gaze only.

| Comp. | Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|---|
| | (Intercept) | −2.045630 | 0.315946 | −6.47 | *** |
| c2 | e1 | −0.041060 | 0.438268 | −0.09 | |
| | l1 | 0.016237 | 0.247976 | 0.07 | |
| | l2 | 0.643117 | 0.160811 | 4.00 | *** |
| c2:l1 | e1:l1 | −0.020054 | 0.313261 | −0.06 | |
| c2:l2 | e1:l2 | 0.089253 | 0.185973 | 0.48 | |

The column Comp. provides information on which effect in the nested structure the reported effect can be compared to.
( . — $p < 0.1$ , ∗ — $p < 0.05$ , ∗∗ — $p < 0.01$, ∗∗∗ — $p < 0.001$)
(e1 — Expected vs Unexpected , l1 — FvsCP , l2 — CvsP)

**Table B.3**
Fixed effects for PMN time-window in the Noun region — Absent Gaze only.

| Comp. | Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|---|
| | (Intercept) | −1.365710 | 0.330022 | −4.14 | *** |
| c3 | e1 | −1.359080 | 0.495451 | −2.74 | ** |
| | l1 | 0.015517 | 0.199083 | 0.08 | |
| | l2 | 0.470537 | 0.099384 | 4.73 | *** |
| c3:l1 | e1:l1 | −0.380542 | 0.303344 | −1.25 | |
| c3:l2 | e1:l2 | 0.320965 | 0.220242 | 1.46 | |

The column Comp. provides information on which effect in the nested structure the reported effect can be compared to.
( . — $p < 0.1$ , ∗ — $p < 0.05$ , ∗∗ — $p < 0.01$, ∗∗∗ — $p < 0.001$)
(e1 — Expected vs Unexpected , l1 — FvsCP , l2 — CvsP)

**Table B.4**
Fixed effects for N400 time-window in the Noun region — Interaction Model.

| Comp. | Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|---|
| (Intercept) | | −1.350640 | 0.372633 | −3.62 | *** |
| c1 | g1 | 2.184170 | 0.532237 | 4.10 | *** |
| | e1 | −0.703295 | 0.335196 | −2.10 | * |
| l1 | l1 | −0.564405 | 0.381632 | −1.48 | |
| l2 | l2 | 0.717834 | 0.124166 | 5.78 | *** |
| | g1:e1 | −0.067584 | 0.684157 | −0.10 | |
| c1:l1 | g1:l1 | 0.766438 | 0.345136 | 2.22 | * |
| c1:l2 | g1:l2 | −0.053214 | 0.263756 | −0.20 | |
| | e1:l1 | −0.289596 | 0.226858 | −1.28 | |
| | e1:l2 | 0.160414 | 0.176423 | 0.91 | |
| | g1:e1:l1 | −0.851636 | 0.517694 | −1.65 | . |
| | g1:e1:l2 | −0.328216 | 0.297961 | −1.10 | |

The column Comp. provides information on which effect in the nested structure the reported effect can be compared to.
Note that the interaction term g1:e1:l1 only reaches marginal significance. Based on the aforementioned potential issues of the interaction analysis and the hypothesis driven contrasts, we nonetheless perform the split based on gaze-presence.
( . — $p < 0.1$ , ∗ — $p < 0.05$ , ∗∗ — $p < 0.01$, ∗∗∗ — $p < 0.001$)
(g1 — Present vs Absent Gaze , e1 — Expected vs Unexpected , l1 — FvsCP , l2 — CvsP)

**Table B.5**
Fixed effects for N400 time-window in the Noun region — Present Gaze only.

| Comp. | Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|---|
| | (Intercept) | −2.445500 | 0.353885 | −6.91 | *** |
| c2 | e1 | −0.526565 | 0.478518 | −1.10 | |
| | l1 | −0.178876 | 0.309354 | −0.58 | |
| | l2 | 0.749282 | 0.207113 | 3.62 | *** |
| c2:l1 | e1:l1 | 0.133837 | 0.372157 | 0.36 | |
| c2:l2 | e1:l2 | 0.323252 | 0.207900 | 1.55 | |

The column Comp. provides information on which effect in the nested structure the reported effect can be compared to.
( . — $p < 0.1$ , ∗ — $p < 0.05$ , ∗∗ — $p < 0.01$, ∗∗∗ — $p < 0.001$)
(e1 — Expected vs Unexpected , l1 — FvsCP , l2 — CvsP)

**Table B.6**
Fixed effects for N400 time-window in the Noun region — Absent Gaze only.

| Comp. | Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|---|
| | (Intercept) | −0.258060 | 0.542533 | −0.48 | |
| c3 | e1 | −0.595568 | 0.545232 | −1.09 | |
| | l1 | 0.585008 | 0.252534 | 2.32 | * |
| | l2 | 0.695161 | 0.153443 | 4.53 | *** |
| c3:l1 | e1:l1 | −0.705517 | 0.323011 | −2.18 | * |
| c3:l2 | e1:l2 | −0.001267 | 0.245578 | −0.01 | |

The column Comp. provides information on which effect in the nested structure the reported effect can be compared to.
( . — $p < 0.1$ , ∗ — $p < 0.05$ , ∗∗ — $p < 0.01$, ∗∗∗ — $p < 0.001$)
(e1 — Expected vs Unexpected , l1 — FvsCP , l2 — CvsP)

**Table B.7**
Fixed effects for P600 time-window in the Noun region — Interaction Model.

| Comp. | Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|---|
| (Intercept) | | −0.166997 | 0.387582 | −0.43 | |
| c1 | g1 | 3.436550 | 0.572052 | 6.01 | *** |
| | e1 | −0.332693 | 0.356325 | −0.93 | |
| l1 | l1 | 0.758392 | 0.263871 | 2.87 | ** |
| l2 | l2 | 0.558162 | 0.234163 | 2.38 | * |
| | g1:e1 | 0.285970 | 0.788716 | 0.36 | |
| c1:l1 | g1:l1 | 1.623840 | 0.476936 | 3.40 | *** |
| c1:l2 | g1:l2 | −0.029773 | 0.274804 | −0.11 | |
| | e1:l1 | 0.217624 | 0.264244 | 0.82 | |
| | e1:l2 | 0.211300 | 0.173714 | 1.22 | |
| | g1:e1:l1 | −1.087500 | 0.542457 | −2.00 | * |
| | g1:e1:l2 | −0.526769 | 0.347419 | −1.52 | |

The column Comp. provides information on which effect in the nested structure the reported effect can be compared to.
Note that while the interaction term g1:e1:l1 reaches significance, none of the follow-up models (Tables B.8 and B.9) confirm this effect beyond a marginal significance for e1:l1 and e1:l2 when gaze was present.
( . — $p < 0.1$ , ∗ — $p < 0.05$ , ∗∗ — $p < 0.01$, ∗∗∗ — $p < 0.001$)
(g1 — Present vs Absent Gaze , e1 — Expected vs Unexpected , l1 — FvsCP , l2 — CvsP)

**Table B.8**
Fixed effects for P600 time-window in the Noun region — Present Gaze only.

| Comp. | Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|---|
| | (Intercept) | −1.886640 | 0.457274 | −4.13 | *** |
| c2 | e1 | −0.480855 | 0.551334 | −0.87 | |
| | l1 | −0.054053 | 0.346524 | −0.16 | |
| | l2 | 0.573873 | 0.277242 | 2.07 | *** |
| c2:l1 | e1:l1 | 0.757047 | 0.401424 | 1.89 | . |
| c2:l2 | e1:l2 | 0.475262 | 0.241977 | 1.96 | . |

The column Comp. provides information on which effect in the nested structure the reported effect can be compared to.
( . — $p < 0.1$ , ∗ — $p < 0.05$ , ∗∗ — $p < 0.01$, ∗∗∗ — $p < 0.001$)
(e1 — Expected vs Unexpected , l1 — FvsCP , l2 — CvsP)

**Table B.9**
Fixed effects for P600 time-window in the Noun region — Absent Gaze only.

| Comp. | Fixed effect | β-estimate | Std. error | t value | sig. |
|---|---|---|---|---|---|
| | (Intercept) | 1.552080 | 0.505207 | 3.07 | ** |
| c3 | e1 | −0.199031 | 0.511507 | −0.39 | |
| | l1 | 1.576180 | 0.366025 | 4.31 | *** |
| | l2 | 0.542937 | 0.265304 | 2.05 | * |
| c3:l1 | e1:l1 | −0.327191 | 0.355525 | −0.92 | |
| c3:l2 | e1:l2 | −0.054001 | 0.249057 | −0.22 | |

The column Comp. provides information on which effect in the nested structure the reported effect can be compared to.
( . — $p < 0.1$ , ∗ — $p < 0.05$ , ∗∗ — $p < 0.01$, ∗∗∗ — $p < 0.001$)
(e1 — Expected vs Unexpected , l1 — FvsCP , l2 — CvsP)

## References


Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

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

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge U Press.

Baron-Cohen, S. (1997). How to build a baby that can read minds: Cognitive mechanisms in mindreading. *The Maladapted Mind: Classic Readings in Evolutionary Psychopathology*, 207–239.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bayliss, A. P., Frischen, A., Fenske, M. J., & Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition*, *104*(3), 644–653.

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic Bulletin & Review*, *13*(6), 1061–1066.

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences*, [ISSN: 13646613] *12*(7), 254–258.

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Observing shared attention modulates gaze following. *Cognition*, *120*, 292–298.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*(4), 787.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., & Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, *117*(3), 133–152.

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.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, *98*(2), 159–168.

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, *9*(1), 55–72.

Chen, Y.-C., & Yeh, S.-L. (2012). Look into my eyes and I will see you: Unconscious processing of human gaze. *Consciousness and Cognition*, *21*(4), 1703–1710.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256–266.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*(1), 21–58.

Degen, J., & Tanenhaus, M. K. (2019). Constraint-based pragmatic processing. *The Oxford Handbook of Experimental Semantics and Pragmatics*, 21–38.

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.

Donchin, E. (1981). Surprise!...surprise? *Psychophysiology*, *18*(5), 493–513.

Driver IV, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, *6*(5), 509–540.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy*, *5*(1), 39–60.

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.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495,

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 319.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *4*(11), 274–279.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia*, *38*(11), 1518–1530.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615,

Hietanen, J. K., Leppänen, J. M., Nummenmaa, L., & Astikainen, P. (2008). Visuospatial attention shifts by gaze and arrow cues: An ERP study. *Brain Research*, *1215*, 123–136,

Jachmann, T. K. (2020). *The immediate influence of speaker gaze on situated speech comprehension: evidence from multiple ERP components*. Saarländische Universitäts-und Landesbibliothek, http://dx.doi.org/10.22028/D291-31309.

Jachmann, T. K., Drenhaus, H., Staudte, M., & Crocker, M. W. (2019). Influence of speakers' gaze on situated language comprehension: Evidence from event-related potentials. *Brain and Cognition*, *135*, Article 103571.

Jachmann, T. K., Drenhaus, H., Staudte, M., & Crocker, M. W. (2022). When a look is enough: Neurophysiological correlates of referential speaker gaze in situated comprehension (data). https://osf.io/uzktp/.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156.

Kappenman, E. S., & Luck, S. J. (2016). Best practices for event-related potential research in clinical populations. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(2), 110–115.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481–529.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127.

Kreysa, H. (2009). *Coordinating speech-related eye movements between comprehension and production*. The University of Edinburgh.

Kreysa, H., Nunnemann, E. M., & Knoeferle, P. (2018). Distinct effects of different visual cues on sentence comprehension and later recall: The case of speaker gaze versus depicted actions. *Acta Psychologica*, *188*, 220–229.

Kuhn, G., & Kingstone, A. (2009). Look away! eyes and arrows engage oculomotor responses automatically. *Attention, Perception, & Psychophysics*, *71*(2), 314–327.

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, *32*(1), 12–35.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.

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

MATLAB (2017). *version 9.2.0 (R2017a)*. Natick, Massachusetts: The MathWorks Inc..

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects : eye movements during noun phrase production. *Cognition*, *66*, 25–33.

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development*, *21*(3), 469–482.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience*, *4*(1), 15–22.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148.

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review*, *9*(3), 507–513.

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39.

Sekicki, M., & Staudte, M. (2018). Eye'll help you out! how the gaze cue reduces the cognitive load required for reference processing. *Cognitive Science*, *42*(8), 2418–2458.

Sharbrough, F., Chatrian, G. E., Lesser, R. P., Lüders, H., Nuwer, M., & Picton, T. W. (1991). American electroencephalographic society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *8*, 200–202.

Staudte, M., & Crocker, M. W. (2010). When robot gaze helps human listeners: Attentional versus intentional account. In *Proceedings of the 32nd annual meeting of the cognitive science society* (pp. 1637–1642).

Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, *120*(2), 268–291,

Staudte, M., Crocker, M. W., Heloir, A., & Kipp, M. (2014). The influence of speaker gaze on listener comprehension: Contrasting visual versus intentional accounts. *Cognition*, [ISSN: 18737838] *133*(1), 317–328.

Stein, T., Senju, A., Peelen, M. V., & Sterzer, P. (2011). Eye contact facilitates awareness of faces during interocular suppression. *Cognition*, *119*(2), 307–311.

Symons, L. A., Hains, S. M., & Muir, D. W. (1998). Look at me: Five-month-old infants' sensitivity to very small deviations in eye-gaze during social interactions. *Infant Behavior and Development*, *21*(3), 531–536.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Vol. 268*, In *Science (New York, N.Y.)* (5217), (pp. 1632–1634).

Thomsen, C. E. (1974). Eye contact by non-human primates toward a human observer. *Animal Behaviour*, *22*(1), 144–149.

Tomasello, M., Call, J., & Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Animal Behaviour*, *55*(4), 1063–1069.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 1454–1463.

Tomasello, M., & Kruger, A. C. (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language*, *19*(2), 311–333.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134.

Uono, S., Sato, W., & Kochiyama, T. (2014). Commonalities and differences in the spatiotemporal neural dynamics associated with automatic attentional shifts induced by gaze and arrows. *Neuroscience Research*, [ISSN: 18728111] *87*(C), 56–65.

Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology*, *34*(2), 131–156.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162.