Cognition 214 (2021) 104744

Contents lists available at ScienceDirect

Cognition

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

The state of the onion: Grammatical aspect modulates object representation during event comprehension

Julia Misersky [a,b,*], Ksenija Slivac [a,b], Peter Hagoort [a,c], Monique Flecken [a,d]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[b] International Max Planck Research School for Language Sciences, the Netherlands
[c] Donders Institute, Radboud University Nijmegen, the Netherlands
[d] University of Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Event cognition
Language comprehension
Grammatical aspect
Event-related potentials

ABSTRACT

The present ERP study assessed whether grammatical aspect is used as a cue in online event comprehension, in particular when reading about events in which an object is visually changed. While perfective aspect cues holistic event representations, including an event's endpoint, progressive aspect highlights intermediate phases of an event. In a 2 × 3 design, participants read SVO sentences describing a change-of-state event (e.g., to chop an onion), with grammatical Aspect manipulated (perfective "*chopped*" vs progressive "*was chopping*"). Thereafter, they saw a Picture of an object either having undergone substantial state-change (SC; a chopped onion), no state-change (NSC; an onion in its original state) or an unrelated object (U; a cactus, acting as control condition). Their task was to decide whether the object in the Picture was mentioned in the sentence. We focused on N400 modulation, with ERPs time-locked to picture onset. U pictures elicited an N400 response as expected, suggesting detection of categorical mismatches in object type. For SC and NSC pictures, a whole-head follow-up analysis revealed a P300, implying people were engaged in detailed evaluation of pictures of matching objects. SC pictures received most positive responses overall. Crucially, there was an interaction of Aspect and Picture: SC pictures resulted in a higher amplitude P300 after sentences in the perfective compared to the progressive. Thus, while the perfective cued for a holistic event representation, including the resultant state of the affected object (i.e., the chopped onion) constraining object representations online, the progressive defocused event completion and object-state change. Grammatical aspect thus guided online event comprehension by cueing the visual representation(s) of an object's state.

1. Introduction

Understanding the events happening in the world around us is fundamental to our day-to-day lives. The research field of *event cognition* is concerned with how we perceive and understand events in the interface with action, vision, memory, and language (e.g., Zacks, 2019). Specific to characterisations of events is the dimension of time during which an event unfolds, as well as the specific relationship between the people and objects in an event. In this project, we explore how the expression of time in language may modulate how we understand events online. Specifically, we are studying whether grammatical aspect marking (perfective vs progressive) affects mental representations of objects that are being affected, and their states, in events.

An event can be described as 'a segment of time at a given location that is conceived by an observer to have a beginning and an end' (Zacks & Tversky, 2001, p. 3), highlighting time as a crucial defining feature of an event. The segmentation of our continuous experiences into events with discrete boundaries is afforded by predictive processes (event segmentation theory (EST); Zacks, Speer, Swallow, Braver, & Reynolds, 2007, Zacks & Tversky, 2001). To support event comprehension and segmentation, we rely on our prior knowledge of similar types of events, including the specific relations between event entities like people and objects, encapsulated in memory representations (*event models*; Richmond & Zacks, 2017). Recently, it has been proposed that objects and their affordances are particularly important in guiding event comprehension. Altmann and Ekves's Intersecting Object Histories theory (IOH, 2019) focuses on objects in events as key to understanding an event's inner (temporal) structure. In IOH, an event is defined by one or more

* Corresponding author at: P.O. Box 310, 6500 AH Nijmegen, the Netherlands.
E-mail addresses: julia.misersky@mpi.nl, m.e.p.flecken@uva.nl (J. Misersky).

https://doi.org/10.1016/j.cognition.2021.104744
Received 24 January 2020; Received in revised form 30 March 2021; Accepted 21 April 2021
Available online 4 May 2021
0010-0277/© 2021 Elsevier B.V. All rights reserved.

object(s) undergoing change across time and space. In their example of an event like "*The chef chopped the onion.*", the onion undergoing change (i.e., it is being chopped) and its relation to the chef, as well as the instruments used to create the change over time, is what Altmann and Ekves call *ensembles of intersecting object histories*. To understand an event, then, requires a person to track different object states while the event unfolds. This means that other entities within the event structure, including for example the chef who has the role of the agent in the onion chopping event, are understood through tracking of object histories. This happens through relational binding, a process tightly linked to our prior experiences, by which context and time allow for the tokenization of the object and other event entities. Altmann and Ekves explain that as a result 'object tokens essentially carry their history with them and are in fact unique trajectories through representational space and time.' (2019, p. 28). In other words, keeping track of objects and their states has spatio-temporal contingencies, which allow us to understand the transition an object goes through during the time course of an event by tracking its change in state (e.g., an onion going from being whole to being chopped up into pieces). In line with this, fMRI results suggest that multiple object states (for example, an object's original state, a whole onion, as well as its resultant state, a chopped onion) are represented, and compete with one another, when understanding change-of-state events through language (Hindy, Altmann, Kalenik, & Thompson-Schill, 2012; Solomon, Hindy, Altmann, & Thompson-Schill, 2015). In sum, keeping track of how, and to what degree, objects change over time as an event unfolds, is key to event comprehension.

Importantly, we can understand events in two ways: Directly, through experience, and indirectly through language (Altmann & Ekves, 2019; Radvansky & Zacks, 2014). The ways in which time is encoded linguistically, however, differs across languages (Klein, 2004). For example, some well-studied languages such as English express temporal information grammatically through tense and aspect marking on the verb (Klein, 1994). Simplified it can be said that verb tense specifies when the situation described happened relative to utterance time, the time at which the sentence was uttered. A past tense thus locates an event at a time interval before the moment of speaking. This holds for English simple past tense (e.g., *chopped*) as well as the past progressive (e.g., *was chopping*). Grammatical aspect marks different viewpoints on the internal temporal structure of the same event, such as completed or ongoing. This means that aspect provides information as to what time interval or alternatively which *phase* of the entire situation is under discussion (Klein, 1994); it specifies the so-called Topic Time, 'the time for which the particular utterance makes an assertion' (Klein, 1994, p. 37). Perfective aspect asserts that the event as a whole took place from beginning to end, while imperfective aspect asserts that a part of the event took place, without committing that it actually reached its endpoint (e.g., the event may have been interrupted).

As a result, tense and aspect function independent from one another. Take the example of *The chef chopped the onion.* versus *The chef was chopping the onion*. In both sentences, the verb tense situates the Topic Time of the event at some point in the past. In the first sentence the simple past tense, in this case signalling perfective aspect in English, marks the situation as completed, i.e., the Topic Time thus covers the entire situation, rendering a holistic viewpoint on the event. In contrast, the second sentence with progressive aspect marks ongoingness, stating that there was an interval of onion chopping going on at the Topic Time. The progressive thus highlights an intermediate phase of the event and defocusses its boundaries.

Functions and definitions of tense, aspect and combinations of the two pose a matter of discussion, for example when verbs describe states (Klein, 2004). However, verb predicates that refer to actions with an inherent endpoint (so-called 'telic' predicates, e.g., to chop an onion) in the past tense, can distinctly elicit interpretations of ongoing (past progressive) and completed events (simple past) in English (see e.g., Smith, 1991). Aspect thus allows us to view an event from distinct viewpoints, either emphasising (perfective) or diffusing (progressive) event boundaries.

Grammatical aspect can affect the construction of an event model and, because of this, event segmentation (Feller, Eerland, Ferretti, & Magliano, 2019). Specifically, aspect provides detailed information about the temporal contours of a situation, and can thus give specific cues as to how the time interval under discussion is related to the event's boundaries, in case of the perfective-progressive distinction specifically, the *right* boundary (the point in time at which a given event ends, its endpoint) is at stake. Importantly, aspect marking may be particularly relevant to events during which an object undergoes change: it gives us detailed information about the event content by honing in on the inner temporal structure of an event (e.g., Ferretti, Kutas, & McRae, 2007; Zwaan, 2008), expressing explicitly how far along the trajectory of change the event has progressed −that is, how far the event has advanced towards reaching its endpoint reflected in a substantial change of state in an object− with consequences for what we imagine the object to look like.

In psycholinguistic research, the sentence-picture verification task has been used to test how language more generally may guide event comprehension, including how we visually represent event time and entities, such as objects. During the sentence-picture verification task, participants read a sentence, which is followed by a picture. Their task is to decide whether the picture shows what was described in the sentence. Along the lines of a match-mismatch task, participants are faster and more accurate in responding to pictures they perceive to match the previously presented sentence, compared to when they perceive a mismatch between picture and sentence.

Madden and Zwaan (2003) used this paradigm to study the effects of aspect on event comprehension. Specifically, their focus was on sentences describing accomplishments (Vendler, 1967), which refer to durative events with an inherent endpoint (e.g., to make a fire). In Experiment 2, they employed a sentence-picture verification task and presented participants with sentences in the perfective (made a fire) or progressive aspect (was making a fire). Each sentence was followed by a drawn picture, which either showed the event in progress or at completion. Participants had to indicate whether the picture matched the sentence. In line with theoretical considerations about grammatical aspect, the authors considered the completed picture to match sentences in the perfective, and in-progress pictures to match sentences in the progressive. Participants were faster to respond to matching pictures compared to mismatching ones after reading a sentence in the perfective. There was no such match benefit effect for pictures following sentences in the progressive; participants responded equally fast to both matching and mismatching pictures. Madden and Zwaan (2003) interpret their results as participants having used the aspectual cues provided by the sentences to build their event model and guide their responses. Crucially, the type of information contained within the event model differed depending on the aspectual cue provided. The perfective constrained the event model such that the focus was on the resultant state, leaving only the event endpoint picture considered as a match. The progressive, however, foregrounded features relevant to the intermediate phases of the event, meaning there were fewer constraints and both pictures could be considered a (mis)match.

Aspect thus has the power to cue for a specific event phase, and reaction-time data has shown that aspect also influences how we think about instruments in events (Madden-Lombardi, Dominey, & Ventre-Dominey, 2017). French native speaking participants read sentences (perfective vs. imperfective), where the critical instrument (in-use vs not-in-use) as well as the target objects (ongoing vs resultant state) were replaced by images (e.g., *John was using/had used a corkscrew to open the bottle.*). In the match condition both the instrument and the object pictures matched the temporal constraints of the sentence (e.g., reading *John had used…,* and seeing the corkscrew in use as well as the bottle opened, all signalling a completed event). In addition, there were trials where only the instrument or the object picture matched (e.g., reading *John had used…,* and seeing the corkscrew in use as but the bottle half closed, with only sentence and corkscrew signalling a complete action). The authors found a match effect, such that, when aspect and the pictures matched, participants were faster to respond to the question of whether the sentence was acceptable. The match effects were strongest for objects in perfective sentences compared to instruments. Overall, the match effect for objects in perfective sentences highlights that perfective aspect focuses strongly on the endpoint (right boundary) of the event.

This type of match-mismatch effect has also been studied using magnetoencephalogram (MEG) and electroencephalogram (EEG). Hirschfeld, Zwitserlood, and Dobel (2011) presented participants with a spoken sentence suggesting the shape of an object ("The ranger saw a duck in the lake." vs. "The ranger saw a duck in the air.") or an unrelated sentence ("The ranger prepared a sandwich."), followed by a picture of the object in one of two states (sitting vs. flying duck). Using MEG, the authors measured M1 and M400 amplitude from picture onset. The M1 (here: 115–140 milliseconds) represents early processing of perceptual information. Hirschfeld et al. (2011) found a M1 modulation, with the matching shape condition eliciting stronger responses compared to the other conditions, which they interpreted as a processing advantage. The M400 (here: 300–450 milliseconds) is related to the processing of semantics. It is modulated by the perceived mismatch of semantic content, with a higher amplitude relating to a higher degree of perceived mismatch. In the cognition of action events, the N400 has been reliably found for unpredicted event happenings (Amoruso et al., 2013; Zacks, 2019). Interestingly, following either sentence suggesting object shape, no N400 effect was observed in Hirschfeld et al. (2011). However, the authors observed an N400 effect for unrelated sentences. Similar N400 results were obtained in an EEG study (Hirschfeld, Feldker, & Zwitserlood, 2012).

Using similar stimuli as Hirschfeld et al. (2011, 2012), Zwaan and Pecher (2012) aimed to replicate six behavioural studies using an online version of the sentence-picture verification task. In the tasks, an English sentence suggested either the orientation ("*John put the pencil in the drawer/in the cup.*", implying horizontal or vertical orientation respectively; Stanfield & Zwaan, 2001), the shape (an egg in a carton/in a pan; Zwaan, Stanfield, & Yaxley, 2002) or the colour ("*Sarah stopped in the woods to pick a leaf off the ground/the tree.*"; Connell, 2007) of an object. The picture either matched or mismatched the suggestion. The results indeed replicate a match-mismatch effect for orientation and shape, with the effect for shape being particularly strong. There was no replication of the match effect for colour. For both orientation and shape, the sentence guided the participants' responses to the objects in the pictures. Zwaan and Pecher (2012) consider the match advantage to reflect an automatic mental simulation process in part afforded by relevance (here: of orientation and shape respectively), that can also be observed in other cognitive tasks. While some situate the match-mismatch effects of the sentence-picture verification task within mental simulation theory (Zwaan et al., 2002; Zwaan & Pecher, 2012), recent research has challenged this interpretation (Ostarek, Joosen, Ishag, De Nijs, & Huettig, 2019). Note, that the aim of the current study was not to test mental simulation theory but rather to assess the role of time in language, in event comprehension with regards to event cognition theory.

Recently, behavioural research using the sentence-picture verification task has made a step into this direction and assessed the effect of language cues on object-state representation (Kang, Eerland, Joergensen, Zwaan, & Altmann, 2019). In Experiments 2 (past tense sentences) and 3 (future tense sentences), participants were presented with a sentence context either suggesting an object undergoing substantial state-change ("*The woman dropped/will drop the ice cream*") or minimal state-change ("*The women chose/will choose the ice cream*"). The sentence was followed by a picture of the object in the original, unchanged state or a resultant, changed state. The results suggest a match effect: For both past and future tense sentences, the modified pictures were responded to faster when the sentence suggested a state-change. The match effect for original, unchanged pictures was only found after past tense sentences. The authors highlight two limitations of their research. First, the behavioural nature of the study limits the conclusion to processes relevant to explicit decision-making. Second, the degree of change was manipulated using different verbs (e.g., "*drop*" vs "*choose*"), which may give rise to different semantic associations between the actions denoted by the verbs and the objects as depicted in the pictures, e.g., a picture of a dropped ice cream is more strongly associated with the substantial change verb 'drop' than the minimal change verb 'choose' (cf. Kang et al., 2019). Despite this, the results highlight that object states can be activated through the use of verbs as suggested also by IOH (Altmann & Ekves, 2019), and that the state of an object is indeed part of event comprehension and the mental images we construct during this process.

In addition to the above example studying the effect of tense on event comprehension, there is some evidence that aspect, too, has the general potential to constrain event interpretation, at least in behavioural research. In contrast to the perfective, progressive aspect leads participants to rate short events to be longer, and long events to be shorter (Flecken & Gerwien, 2013), and can affect participants' perception of event boundaries such that accomplishments (Becker, Ferretti, & Madden-Lombardi, 2013; Magliano & Schleich, 2000) and motion events (Anderson, Matlock, & Spivey, 2013; Athanasopoulos & Bylund, 2013) are perceived as ongoing in certain circumstances. In sum, the sentence-picture verification task has been used to study how we represent objects in relation to language. In particular, the experimental manipulation of verb aspect has shown that the perfective renders a focus on event endpoints, whereas the progressive focuses on an event's ongoingness, while defocusing its boundaries.

1.1. *The present study*

The sentence-picture verification task presents a suitable paradigm to focus on object state change within events, such that we can get a clearer understanding of how grammatical aspect can affect how we represent object states in events. In particular, we can study whether aspect cues a particular phase of the event thereby modulating the expected instantiations of an object undergoing change. This will allow us to see how empirical data scales up to recent theoretical approaches in event cognition, not focussing on the segmentation of events from a stream of activity, but on event content and inner (temporal) structure (see IOH; Altmann & Ekves, 2019). In doing so, the current EEG study will specifically shed light onto the role of language as an indirect means to experience events and how this affects event understanding more generally.

Extending previous research (e.g., Hindy et al., 2012; Kang et al., 2019) in the domain of verb semantics to the realms of grammatical aspect marking, we focused specifically on change-of-state events. We presented participants with sentences, in which *all* verbal predicates suggested a change-of-state in an object (e.g., to chop an onion). In addition, we wanted to ensure that at the beginning of the event described, the object would cue a specific visual representation (i.e., it exists in an original state), and that visual representations of right and left boundary of the event were not interchangeable (as with e.g., the verbs to open/close). To that end, we used verbs in which the change suggested was irreversible (e.g., to burn), and avoided verbs of creation (e.g., to build). We manipulated the degree of change expressed by the sentence, by presenting the verb phrase in the perfective aspect (simple past tense; *chopped*) or in the (past) progressive (*was chopping*). All sentences were followed by either a picture of the object having undergone a state-change (SC; a chopped onion) or no state-change (NSC; an onion in its original, unchopped state, cf. Kang et al., 2019) or unrelated objects (U; a cactus). Including the unrelated condition addressed a limitation of previous work (Kang et al., 2019), and allowed us to establish a baseline of semantic mismatch (see also below). Different from related studies on aspect and event comprehension, we did not include images reflecting an ongoing event phase, such as a half-chopped onion. Our primary focus however, was event boundaries. Specifically, we looked at the respective focusing or defocusing of the event's right boundary (the endpoint of the event, characterized by a substantially changed object), through perfective/progressive marking on verbs with change of state semantics. In order to build on studies framed in the context of the IOH theory (Altmann & Ekves, 2019), we used stimuli, which characterized both the objects' original (NSC) and their changed, resultant states (SC) (cf. Hindy et al., 2012; Kang et al., 2019; Solomon et al., 2015).

The participants' task was to indicate whether the object in the picture was mentioned in the sentence. This setup allowed us to measure whether participants were reading attentively, whilst also keeping them occupied throughout the experiment. Different from previous work on aspect and event comprehension, but similar to studies using the sentence-picture verification task, our manipulation of object state was thus task- irrelevant (U pictures should elicit No responses, while both SC and NSC pictures should receive a Yes). Further, a response was not required on each trial, as the focus of this study was not on people's behavioural responses, but on the ERPs elicited by the pictures. Our interest in implicit differential processing of the two sentence-related object pictures (SC/NSC pictures of the same object) allowed us to move beyond post-hoc introspective judgements. This way, we were able to explore the more subtle effect aspect may have on how events are understood as they unfold, and to shed light on the underlying cognitive processes involved. Unlike previous research looking at aspect and event cognition (Madden & Zwaan, 2003), we focussed our setup and analysis on how processing of the *same* object state is modulated by aspect. Firstly, this allowed us to study if the object-state as represented by the SC pictures (i.e., the object's changed, or resultant state) is less available after the progressive compared to the perfective. Secondly, we were able to assess whether the state depicted by NSC pictures (objects in their canonical, original states) is more available after the perfective, given the holistic event viewpoint of the situation in its entirety.

We recorded ERPs from picture onset and first focused on N400 amplitude (300 to 500 milliseconds) to examine how grammatical aspect would modulate participants' conceptualisation of the events described as characterized by the pictures of the objects in a specific state. We took the N400 as indicative of the degree of match between sentence and visual stimulus, with more negative N400 amplitudes reflecting lower degree of match. NSC pictures depicted the object at beginning of the event, and SC pictures showed its resultant state at the end. NSC and SC pictures thus represented the boundaries of the event. U pictures served as a control. Given this, we expected U pictures to result in a high amplitude N400 indicative of a semantic mismatch (Kutas & Hillyard, 1980) and in the behavioural responses, a No. Further, we hypothesised that overall SC pictures would be considered more compatible with our sentences given that all verbs described a state change event in the past, with aspect manipulated. We thus expected a main effect of picture, such that SC pictures should yield a smaller N400 overall compared to NSC pictures.

Crucially, in the ERP data we expected Aspect and Picture type to interact, i.e., we expected the same Pictures to be processed differently based on the Aspect manipulation. Recall that the perfective has previously been found to cue for an event holistically including its completion phase. In contrast, the progressive tends to defocus event boundaries by zooming in on intermediate phases also for events in the past. A representation of an object substantially changed in state should thus be less available after reading a progressive sentence compared to perfective sentences — despite the focus on state-change, as expressed by the semantics of the verbs. So, the initial state of an object (represented by NSC pictures) as well as its resultant state (represented by SC pictures) are in principle not part of the intermediate, event representation evoked by progressive sentences, which instead provides only a snapshot of the ongoing situation between the event boundaries. This is different from the holistic viewpoint elicited by the perfective, which might leave beginning *and* end state (represented by NSC and SC pictures, respectively) available (see also Hindy et al., 2012 for effects of competition between the two states after reading simple past tense [perfective] sentences). However, given the overall strong compatibility of SC pictures with *all* sentences describing a state-change, and the expected main effect of Picture, we expected an interaction of Picture and Aspect. Specifically, SC pictures were expected to elicit a smaller N400 after a sentence in the perfective ("John chopped the onion."), as compared to a sentence in the (past) progressive ("John was chopping the onion."), whilst any potential difference between aspect conditions was expected to be smaller, or even non-existent, for NSC pictures. The compatibility of SC pictures with all sentences was expected to attenuate the degree of match between NSC pictures and the perfective, meaning both Aspect conditions would elicit comparable N400 responses.

2. Methods

2.1. Sentence pre-test

2.1.1. Participants

Eighty-four native speakers of English (48 female, 36 male, 18–39 years of age, $M = 24.13$, $SD = 5.1$), participated in this experiment. Only 82 participants completed the State Change block, whereas 80 additionally completed the Familiarity and Imageability blocks.

The survey was implemented using the Max Planck Institute's Framework for Interactive Experiments (Frinex; Withers, 2017), and participants were recruited using Prolific (www.prolific.co/). They received a monetary compensation for their time (5 GBP per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, The Netherlands.

2.1.2. Materials and design

In this pre-test we tested verb-object compounds (e.g., *chop an onion*). For the experiment proper we were interested in state-change verbs, which described an action which would irreversibly change the physical appearance of an object (e.g., *chop the onion*; $N = 41$). Verbs had one or two syllables, and were no longer than seven letters. The pre-test additionally included verbs which do not visibly change the object (e.g., *collect the onion*; $N = 41$). Note, however, that in line with our study design (including only change-of-state verbs) these were included only to compare the critical state-change verb against, though future use in a follow-up study is possible. Additionally, there was a set of filler items (e.g, *cuddle the teddy bear*; $N = 41$). Each verb could go with one of two objects, resulting in a total of 246 items. All items were split into two lists of 123 items each, which were randomised for each participant.

2.1.3. Procedure

In a block design, participants rated all items on three dimensions: in the first block, items were rated for *State Change*, in the second block for *Familiarity* and in the third block for *Imageability* (cf. Hindy et al., 2012). Items within each block were randomised. For State Change, participants were asked to what degree the object is changed visibly by the action described, which they had to rate on a Likert scale ranging from 1 (= not changed at all) to 7 (= changed substantially). For Familiarity, participants had to indicate whether they are familiar with the action on a Likert scale ranging from 1 (= not familiar at all) to 7 (= very familiar). For Imageability, participants were asked whether they can create a clear mental image of the situation, which they had to rate on a Likert scale ranging from 1 (= no clear image at all) to 7 (= very clear image).

2.1.4. Data analysis and results

Using the lmerTest package (Version 3.0–0; Kuznetsova, Brockhoff, & Christensen, 2017) in R (Version 3.4.2), we analysed the descriptives for participants' ratings of State Change, comparing the state-change verbs ($M = 4.85$, $SD = 1.79$) to the no-change verbs ($M = 2.18$, $SD = 1.79$). We ran a linear mixed effect model in order to establish how verbs that suggest a state change in the object compared to those that do not. *t*-tests used the Satterthwaite's method. Verbs suggesting a state change received significantly higher ratings than those that did not (Intercept: β

= 2.176, $t(122.710) = 20.74$, $p < 0.001$; SC items: $\beta = 2.673$, $t(172.820) = 37.33$, $p < 0.001$), meaning they described more substantial state change. In addition, we obtained overall high ratings of Familiarity (M = 5.29, SD = 2.02), as well as Imageability (*M* = 5.90, *SD* = 1.71). For completeness, we also add the ratings for no-change verbs here. We obtained high ratings for Familiarity (M = 4.53, SD = 2.30), as well as Imageability (*M* = 4.92, *SD* = 2.20) for those verbs. Seven state change verbs were excluded based on mean ratings below 3.5 for State Change. Thirty-four verbs remained for the experiment proper (see list in the Appendix).

2.2. Picture pre-test

2.2.1. Participants

Forty-one native speakers of English (23 female, 1 No Answer, 17 male, 18–34 years of age, *M* = 22.88, *SD* = 4.42), participated in this experiment. The survey was implemented using Limesurvey (www.limesurvey.org), and participants were recruited using Prolific (www.prolific.co/). They received a monetary compensation for their time (5 GBP per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, The Netherlands.

2.2.2. Materials and design

Based on the sentences, multiple colour photographs were taken. Pictures showed familiar household objects having undergone state-change (SC, *N* = 88), no state-change (NSC, *N* = 88) or items for the unrelated condition (U, *N* = 132). Objects were photographed against a white background. The best quality photographs were edited in Adobe Photoshop CS6 (Version 13.0, 2012) such that the sharpness, contrast, lighting and size were similar across all pictures, and the background was removed. A total of 306 pictures were chosen for this pre-test (freely available for download via the MPI for Psycholinguistics Language Archive https://hdl.handle.net/1839/6d395627-9c37-42a8-be90-f5bddc57867b).

2.2.3. Procedure

There were two lists, resulting in each participant rating half of all 306 items (*N* = 153). Per trial, participants saw one picture, which was accompanied by a description. For original-state items, the descriptor was simply the noun (e.g., "*an onion*"). Objects in the resultant state were described using an adverbial (e.g., "*a chopped onion*"). Participants were asked to indicate how well the description fits what is depicted in the picture on a Likert scale ranging from 1 (= does not fit at all) to 5 (= fits very well). Item presentation within each list was randomised for each participant.

2.2.4. Data analysis and results

Participants gave the pictures high ratings overall, on average 4.51 for SC pictures (*SD* = 0.54), 4.61 for U pictures (*SD* = 0.56) and 4.64 for NSC pictures (*SD* = 0.58). Pictures with mean ratings below 3 (*N* = 7) were excluded. In correspondence with the ratings for the sentences, 68 unique objects were used in their SC and NSC picture versions in the experiment proper. In addition, 85 U pictures remained, which were assigned pseudo-randomly (see below).

2.3. Experiment

2.3.1. Participants

Based on previous ERP research using the sentence-picture verification task with 18–20 participants (Hirschfeld et al., 2012, 2011), we aimed to collect at least 30 complete datasets. Thirty-six native speakers of English (predominantly monolingual, with some knowledge of another language), participated in this experiment. Five participants were excluded from further analysis.[1] The data of 31 participants (20 female, 11 male, 18–35 years of age, *M* = 24.1, *SD* = 3.69) remained for further analysis.

Participants were invited via the online participant database of the Max Planck Institute for Psycholinguistics as well as through advertising. They gave written informed consent prior to their participation and received a monetary compensation for their time (8 Euros per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, The Netherlands.

2.3.2. Materials and design

The materials consisted of short sentences followed by colour photographs. All sentences followed the same subject-verb-object (SVO) structure. The subject was either a noun (e.g., *The woman*; *N* = 24), a first name (e.g., *John*, *N* = 90; equal numbers of male and female first names), or a first name in combination with a kinship term (e.g., *Chloe's aunt*; *N* = 90; equal numbers of male and female names and kinship terms; combining of name and kinship term was randomised, adding further variability within the items). The verb described an irreversible action, which was presented in the perfective (simple past) or past progressive (e.g., *chopped/was chopping*). Each verb could go with one of two possible objects (e.g., *the onion./leeks.*). Each sentence was also paired with one of 204 colour photographs of either the object in its resultant state (SC; a chopped onion/a chopped carrot), the object in its original, non-changed state (NSC; a whole onion/a whole carrot) or an unrelated object (U; a cactus; pictures were assigned pseudo-randomly from a list of 85 possible objects).

The study employed a 2 (Aspect: perfective vs. progressive) × 3 (Picture: SC vs. NSC vs. U) design. To account for each of these conditions, participants saw each verb six times, three times in each Aspect condition (perfective or progressive). To avoid participants receiving a verb combined with the same object repeatedly (i.e., seeing an onion four times), each participant received two object types for each of the verbs (e.g. for SC and NSC picture conditions, one participant would receive, *chopped the onion* followed by SC picture, *was chopping the leeks* followed by SC picture, and *chopped the leeks* – NSC picture, *was chopping the onion* – NSC picture). Aspect and Object type combinations were counterbalanced across participants (e.g., the next participant would receive SC pictures preceded by *chopped the leeks* and *was chopping the onion*). U pictures assigned pseudo-randomly for each participant. This counterbalancing procedure resulted in two lists. Presentation of the items was randomised per participant. Additionally, participants were to receive a question on half of the trials. To this end, we set up the trials such that one participant would receive questions on one half of all verbs, and another on the other half. Since our aim with the questions was simply to occupy participants with a task, and have a measure of them reading the sentences attentively, the questions were not part of the counterbalancing procedure.

2.3.3. Procedure

During setup and capping, participants filled in questionnaires regarding handedness and language background. For the experiment proper, participants were seated in a dimly illuminated sound-attenuating testing booth. They were instructed to read sentences attentively, as their task was to answer whether the object in the picture has been mentioned in the previous sentence. These instructions were

[1] Two participants were excluded because of data recording problems, two further participants had only 70% of trials in one or more condition(s) remaining after data pre-processing, and one participant was excluded on the basis of their accuracy on the content questions (see also below).

presented orally by the experimenter, as well as visually on the testing PC. Participants were asked to blink only between sentences or during breaks. Participants could speak to the experimenter using a microphone at any point during the experiment. Experimental materials were presented using the Presentation software (Neurobehavioral Systems, www.neurobs.com). Each sentence was presented using chunk-by-chunk serial visual presentation in the centre of a 24-in. LCD monitor (HP ZR24w). The background was off-white with words presented in black (Lucida console, font size 20). The beginning of each sentence was preceded by a fixation cross (+). Each word was presented for 500 milliseconds with a blank screen of 450 milliseconds between chunks. Sentence-final words were followed by a full stop. After the sentence, participants again saw a blank screen for 450 milliseconds. Thereafter, a colour photograph appeared on the screen for 500 milliseconds, followed by a 500 millisecond blank. On half of the trials, participants then received a Yes/No question ("*Was the object in the picture mentioned in the sentence?*"), which they had to answer via button press. The experiment was split into three blocks of 68 trials. Between blocks, there were self-paced breaks where a drink of water was offered to the participant. See Fig. 1 for an example of the procedure.

2.3.4. EEG data recording

Continuous EEG was recorded from 27 active electrodes (32 electrode set-up, 10–20 system) attached to an elastic cap (actiCAP), with a BrainAmp DC amplifier (Brain Products, Gilching, Germany). The signal was sampled at 500 Hz. One electrode in the cap provided an active ground. Electrooculogram (EOG) was recorded from additional electrodes above and below the eye and at the outer canthi of the eyes. Electrode impedances were kept below 10 kΩ.

The data was pre-processed using the FieldTrip toolbox for EEG/MEG-analysis (www.fieldtriptoolbox.org; Oostenveld, Fries, Maris, & Schoffelen, 2011) in MATLAB. Segments ranging from before 200 ms until 1000 milliseconds after Picture onset were chosen for further analysis. Off-line-filtering included a low-pass filter at 35 Hz and a high-pass filter at 0.1 Hz. Then, an independent component analysis (ICA) was carried out on the continuous data to remove EOG artifacts. The data were then re-referenced to the average of the signal of both mastoids (Luck, 2014), after which the signal of broken/flat channels was repaired by averaging across neighbouring channels. Trials containing signal exceeding ±75 μV were removed. The data were then inspected visually, and trials showing any remaining electrode jumps or drifting were removed. Lastly, a baseline correction was applied in which the signal was normalized relative to a 200-milliseconds Picture-preceding window for display purposes. For statistical analysis, a regression-based baseline correction was computed as part of the statistical model using a 200-millisecond Picture-preceding window (Alday, 2019; see below).

The data sets of two participants were excluded from further analysis, since fewer <70% of all 34 trials remained after pre-processing in at least one of the conditions. This meant that only data sets from participants with at least 24 trials (70.59%) in all conditions remained. The overall average of trials kept for the remaining participants was 31.97 (94.02%). Split by conditions this looked as follows: SC-perfective (M = 31.71, range 25–34), NSC-perfective (M = 32.23, range 27–34), SC-progressive (M = 32, range 24–34), NSC-progressive (M = 32.06, range 26–34), U-perfective (M = 32, range 28–34), U-progressive (M = 31.81, range 26–34). A further data set was excluded due to participants' performance on the content questions that resulted in accuracy at or below chance.

2.3.5. EEG data analysis

ERPs were time-locked to Picture onset. Per participant per trial, the mean amplitude for each channel was extracted for each trial in the 300 to 500 milliseconds time-window after Picture onset (N400; Kutas & Hillyard, 1980). Note that the baseline data consists of the single-trial mean amplitude in the time-window 200 milliseconds prior to stimulus onset. However, waveforms are plotted using traditional subtraction-based baseline correction, though this correction was not applied prior to extraction for the statistical modelling (see below). Analyses on single trial data were run using linear mixed effects models in R (Version 3.4.2) with the lme4 (Version 1.1–21; Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (Version 3.0–0; Kuznetsova et al., 2017) packages. *t*-tests used the Satterthwaite's method.

First, we took a hypothesis driven-approach, in which six electrodes were included reflecting a canonical N400 region (Cz, CP1, CP2, P3/z/4). Second, we also analysed the whole head data. To do this, we grouped all electrodes into clusters defined by two topographical factors (see Fig. 2) designating their Laterality (left, midline, right) and Anteriority (anterior, central, posterior). Sum-coded contrasts for each topographical factor (Anteriority, Laterality) were set up as relative to central-midline electrodes, and both factors were entered as fixed effects into the analyses.

Both the hypothesis-driven and the whole-head analyses followed the same steps when setting up the linear mixed effect model analysis. A

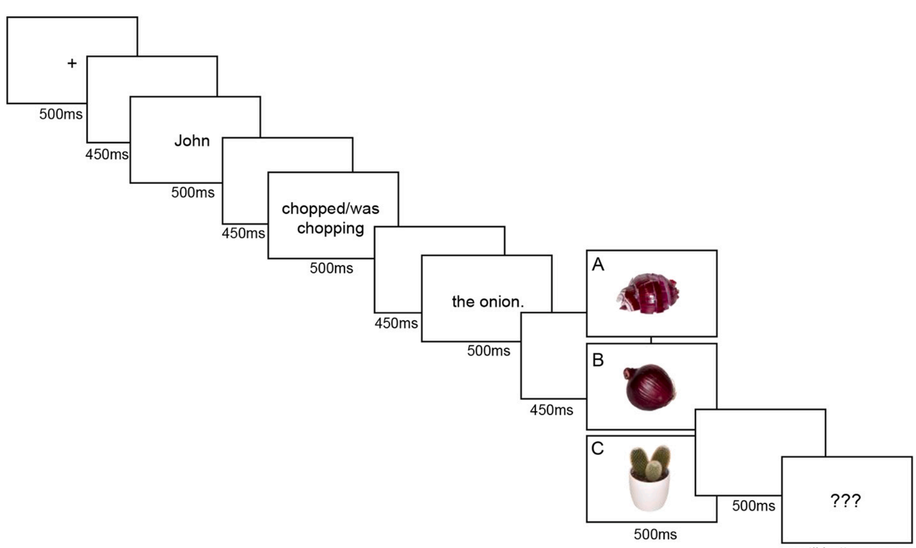

Fig. 1. Procedural EEG setup with all possible Picture manipulations; A showing the state-change (SC), B showing the no state-change (NSC) and C showing the unrelated (U) picture.

	left	midline	right
anterior	F3, F7 FC1, FC5	Fz FCz	F4, F8 FC2, FC6
central	C3 T7	Cz	C4 T8
posterior	CP1, CP5 P3, P7 O1	Pz Oz	CP2, CP6 P4, P8 O2

Fig. 2. Schematic representation of the clusters of electrodes according to the topographical factors of Laterality (left, midline, right) and Anteriority (anterior, central, posterior) for the whole head analysis.

two-step procedure was used to control for individual differences in the data:

In a first step, we created an additional predictor of individual differences for a more robust fixed effects structure in the main analysis (Alday & Kretzschmar, 2019). We did so by fitting a model to the ERP data from the U condition in the 300–500 millisecond time-window consisting only of an intercept and a (centered and scaled) continuous predictor of the pre-stimulus baseline (GLM-based baseline correction, Alday, 2019) in the fixed effects. There is no predictors for condition in this model, as this model was fit from only one condition (i.e., all U pictures). The random effects consisted of by-participant, by-verb and by-object (i.e., the picture participants saw) intercepts. The fitted values from this model reflect the contribution of the pre-stimulus baseline. The residuals (i.e., the ERP response to U pictures not explained by the baseline) thus reflect the baseline-corrected ERP response to U pictures. The residuals were then averaged by participant and trial and then used as a predictor ("calibration") in the second step (see below) as a measure of individual sensitivity to a (canonical) mismatch in object type (between sentence and picture). This step thus allowed for a calibration benchmark of systematic individual variability between participants.

In a second step, we then examined the interaction of Aspect (perfective vs. progressive) and the processing of Picture type (SC vs NSC) as measured by the scaled ERP amplitude in the 300 to 500 millisecond time-window. The fixed effects consisted of (centered and scaled) continuous predictors for pre-stimulus baseline and the predictor derived from the first stage ("calibration"; Alday & Kretzschmar, 2019). We expected the interaction to be driven by the Aspect manipulation, which we implemented in the fixed effects structure by fitting a model with the interaction between Picture and Aspect, but no main effect of Aspect.

In both the hypothesis-driven as well as the whole head analysis, the random effects structure for this main analysis step was simplified until the models had non-singular fit, and included by-participant, by-verb and by-object (i.e., the picture participants saw) intercepts. The specific models for each analysis are shown in Tables 1 and 2.

3. Results

3.1. Behavioural data

Overall, participants showed a high accuracy in correctly responding Yes to SC ($M = 96{,}39\%$) and NSC pictures ($M = 93.83\%$), and No to U pictures ($M = 98.58\%$). For completeness, we also analysed the reaction times (RTs) for correct Yes/No responses. After removing data points above four seconds, RTs of ±2.5 SD of the mean were removed for the analysis; overall, 105 data-points were removed. Picture was sum-coded

Table 1
N400 ROI analysis: Output of the model on ERP amplitude in the 300–500 millisecond time-window on six selected electrodes (lmer (scale(ERP) ~ 1 + Picture * Aspect - Aspect + scale (baseline) + scale (calibration) + (1|participant) + (1|verb) + (1|object).

Fixed Effects	β Estimate	SE	t-Value
Intercept	−0.0045	0.0493	−0.0922
Scale (baseline)	0.2249	0.0142	15.8627***
Scale (calibration)	0.3348	0.0372	8.9980***
Picture [NSC]	−0.0316	0.0141	−2.2465*
Picture [NSC]: Aspect [perfective]	−0.0094	0.0196	−0.4774
Picture [SC]: Aspect [perfective]	0.0310	0.0197	1.5706

*** $p < 0.001$.
* $p < 0.05$.

(SC = 1, NSC = −1, U = 0, and SC = 1, NSC = 0, U = −1). We then examined the interaction of Picture and Aspect (with no main effect of Aspect) as measured by the scaled RTs, including a by-participant random intercept. t-tests used the Satterthwaite's method. The model revealed a main effect of Picture ($\beta = 0.063$, $t(2876) = 2.931$, $p < 0.01$) such that RTs were faster after SC pictures ($M = 815$ ms, $SD = 249$) than after NSC ($M = 855$ ms, $SD = 263$) pictures. There was no interaction between Aspect and Picture. Note that the question was always asked *after* the picture had already disappeared from the screen, and the aim of the questions was to occupy participants with a task, which would give us a reliable measure of them reading the sentences attentively. Further, participants were not urged to respond as quickly as possible upon seeing a question, and there was no timeout of the question. As a result, we will not further discuss these results in regard to our research question.

3.2. Hypothesis-driven ERP data analysis (N400 region of interest)

Table 1 shows the results of the mixed effect regression model on ERP amplitude in the 300 to 500 milliseconds time-window.

Overall, as suggested by the main effect of the "calibration" factor, participants with a higher ERP amplitude response to U pictures, also showed a higher amplitude response to SC and NSC pictures. The model also showed a main effect of Picture type in the 300 to 500 milliseconds time-window: As seen in Fig. 3, SC pictures (solid lines) received the most positive responses ($M = 4.06$, $SD = 4.01$), followed by NSC pictures (dashed lines; $M = 3.26$, $SD = 3.68$). In addition, participants showed an N400 to U pictures following either sentence type (dotted line; $M = -0.12$, $SD = 3.84$), confirming the expected finding of an N400 following unrelated mismatching materials. Unexpectedly, there was no interaction effect for Aspect and Picture. Furthermore, processing of SC and NSC pictures was characterized by a positivity in the same time-window.

3.3. Exploratory ERP data analysis (whole head)

To further investigate the positivity for SC and NSC pictures, we followed-up with a whole-head analysis of all electrodes (grouped into topographical factors) in the 300–500 time-window, following the same steps as the hypothesis-driven analysis.

Main effects of the topographical factors indicate differential activation along the horizontal and lateral axes. In particular, the results for Laterality (left and right) and Anteriority (anterior and posterior), as well as the interaction of the two, point to the positivity being strongest at the central-midline electrodes (see also Fig. 4).

As in the above ROI analysis, there was also a main effect of Picture. As can be seen in Fig. 5, SC pictures (solid lines) again received overall more positive responses compared to NSC pictures (dashed lines). Importantly, Aspect interacted with Picture type: The difference between the solid orange and green lines in Fig. 5 shows that SC pictures following perfective sentences (green solid line) resulted in a more

Table 2
Whole head analysis: Output of the model on ERP amplitude in the 300–500 millisecond time-window (lmer (scale(ERP) ~ 1 + Picture * Aspect * Laterality * Anteriority - Aspect + scale (baseline) + scale (calibration) + (1|participant) + (1|verb) + (1|object).

Fixed Effects	β Estimate	SE	t-Value
Intercept	−0.0046	0.0435	−0.1068
Picture [NSC]	−0.0234	0.0047	−4.9936***
Anteriority [post]	0.2918	0.0065	44.9231***
Anteriority [ant]	−0.2235	0.0065	−34.3994***
Laterality [left]	−0.0282	0.0065	−4.3459***
Laterality [right]	0.0758	0.0065	11.6750***
Scale (baseline)	0.2671	0.0047	57.01500***
Scale (calibration)	0.2684	0.0310	8.6497***
Picture [NSC]:Aspect [perfective]	−0.0094	0.0065	−1.4578
Picture [SC]: Aspect [perfective]	0.0210	0.0065	3.2244**
Picture [NSC]: Anteriority [post]	−0.0049	0.0065	−0.7539
Picture [NSC]: Anteriority [ant]	0.0039	0.0065	0.6010
Aspect [perfective]: Anteriority [post]	−0.0036	0.0065	0.5607
Aspect [perfective]: Anteriority [ant]	0.0040	0.0065	−0.6111
Picture [NSC]: Laterality [left]	−0.0044	0.0065	−0.6829
Picture [NSC]: Laterality [right]	0.0024	0.0065	0.3683
Aspect [perfective]: Laterality [left]	0.0006	0.0065	0.0976
Aspect [perfective]: Laterality [right]	−0.0051	0.0065	−0.7834
Anteriority [post]: Laterality [left]	0.0320	0.0092	3.4806***
Anteriority [ant]: Laterality [left]	−0.0305	0.0092	−3.3267***
Anteriority [post]: Laterality [right]	−0.0103	0.0092	−1.1174
Anteriority [ant]: Laterality [right]	−0.0019	0.0091	−0.2120
Picture[NSC]: Aspect [perfective]: Anteriority [post]	−0.0022	0.0065	−0.3422
Picture [NSC]: Aspect [perfective]: Anteriority [ant]	0.0081	0.0065	1.2524
Picture [NSC]: Aspect [perfective]: Laterality [left]	0.0037	0.0065	0.5743
Picture [NSC]: Aspect [perfective]: Laterality [right]	−0.0007	0.0065	−0.1079
Picture [NSC]: Anteriority [post]: Laterality [left]	−0.0027	0.0092	−0.2918
Picture [NSC]: Anteriority [ant]: Laterality [left]	−0.0020	0.0092	−0.2220
Picture [NSC]: Anteriority [post]: Laterality [right]	−0.00875	0.0092	−0.8165
Picture [NSC]: Anteriority [ant]: Laterality [right]	0.0055	0.0092	0.6041
Aspect [perfective]: Anteriority [post]: Laterality [left]	−0.0014	0.0092	−0.1569
Aspect [perfective]: Anteriority [ant]: Laterality [left]	0.0025	0.0092	0.2751
Aspect [perfective]: Anteriority [post]: Laterality [right]	0.0019	0.0092	0.02065
Aspect [perfective]: Anteriority [ant]: Laterality [right]	0.0012	0.0092	0.1291
Picture [NSC]: Aspect [perfective]: Anteriority [post]: Laterality [left]	−0.0036	0.0092	−0.3889
Picture [NSC]: Aspect [perfective]: Anteriority [ant]: Laterality [left]	0.0013	0.0092	0.1410
Picture [NSC]: Aspect [perfective]: Anteriority [post]: Laterality [right]	0.0037	0.0092	0.4030
Picture [NSC]: Aspect [perfective]: Anteriority [ant]: Laterality [right]	−0.0039	0.0092	−0.4290

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

positive ERP response, compared to SC pictures following progressive sentences (orange solid line). This pattern is also displayed in all panels of Fig. 4, highlighting more positive responses for SC pictures overall, differential procession of SC pictures dependent in aspect. There was no effect of Aspect for NSC pictures. Notably, our factors of interest did not interact with the topographical factors. Thus, the ERP effects for Picture and the interaction of Picture and Aspect highlight a difference in strength between the conditions but not in topographical distribution. In sum, the effects for our factors of interest thus underlie the same ERP component, likely a P300 (see discussion, and waveforms of representative frontal, central and posterior electrodes in Fig. 5).

4. Discussion

The present study assessed whether, and how, grammatical aspect modulates online event comprehension when reading about events, with a focus on (visual) object representation. We used EEG to study at a higher level of granularity the effect of aspect marking on people's mental representation with respect to the specific state of the object (cf. IOH, Altmann & Ekves, 2019). In particular, unlike previous research (e.g., Hindy et al., 2012) we focussed exclusively on change-of-state events and presented people with objects in their original, unchanged state or their changed, resultant state. The goal was to specifically target and shed light onto the role of grammar –here: the expression of event time via aspect marking – as providing important cues to event comprehension.

In line with behavioural research (e.g., Madden-Lombardi et al., 2007; Magliano & Schleich, 2000), we expected perfective aspect to lead to holistic event representations, including the visual instantiation of the state of an object reflecting a completed action, while progressive aspect was expected to lead to a less accessible event completion phase, as reflected in the resultant state of the affected objects. As expected, U pictures, which acted as a control and were designed to be perceived as mismatching the sentence, led to an N400 response, which is in line with research by Hirschfeld et al. (2011, 2012). As revealed by a whole-head follow-up analysis, both SC and NSC pictures elicited a positivity in the 300 to 500 milliseconds time-window along the central-midline, likely a P300. Related to attentional processing, the P300 is often observed in oddball designs and during task-relevant stimulus evaluation (Polich, 2007) and has been linked to evidence accumulation to accurately respond in a task (Kelly & O'Connell, 2015). The P300 consists of two subcomponents, the P3a (frontal) and the P3b (central-parietal). Our effect at central midline electrodes thus resembles a P3b, reflecting detailed processing of task-relevant stimuli, in this case, both SC and NSC pictures. In particular, Kok (2001) proposes that the P3b reflects the matching of an incoming stimulus with a mental representation (a 'template') – here the template was evoked by the sentences– with more positive P3 amplitudes reflecting a higher degree of match. Our finding of a P300 for sentence-related object pictures is in contrast to Hirschfeld et al. (2011, 2012), and may be due to a difference in design. Recall that participants' task was to answer whether the object in the picture was mentioned in the previous sentence. For U pictures, the answer was always No. As such, participants may have gone through a semantic category matching to fulfil the task. While the expected overt responses were the same for SC and NSC pictures (Yes), we show that matching sentence content against objects in (un)changed state differs depending on aspectual cues, reflected in a P300 (P3b) modulation. In order to correctly respond with Yes to SC and NSC pictures, participants had to match the sentence 'template' to two representations of the *same* object. This required them to more closely evaluate the details of SC and NSC pictures, beyond a categorical matching process at the level of object type. In Hirschfeld et al.'s work, however, participants were probed on every trial, whereas our study only asked for sentence-picture verification on half of the trials in a randomised and thus unexpected manner, arguably somewhat similar to an oddball task, which typically triggers P300 responses, due to attentional resources geared towards task-relevant conditions (Polich, 2007).

Our ERP results indicate a main effect of Picture, such that SC objects received a higher amplitude P300 overall, regardless of the Aspect manipulation. NSC pictures, by contrast, received a more attenuated P300 response, again irrespective of the Aspect manipulation. In line with Kok (2001), the larger positivity for SC pictures (changed objects) reflects that these matched the sentences, which all involved state change verbs, more closely than NSC pictures (unchanged objects). Further, our results can be interpreted in line with a so-called 'goal-bias' in cognition: Previous work on action learning (Monroy, Gerson, & Hunnius, 2017), and on the encoding and memory of motion events in infants and young children (Lakusta & Carey, 2015; Lakusta & Landau,

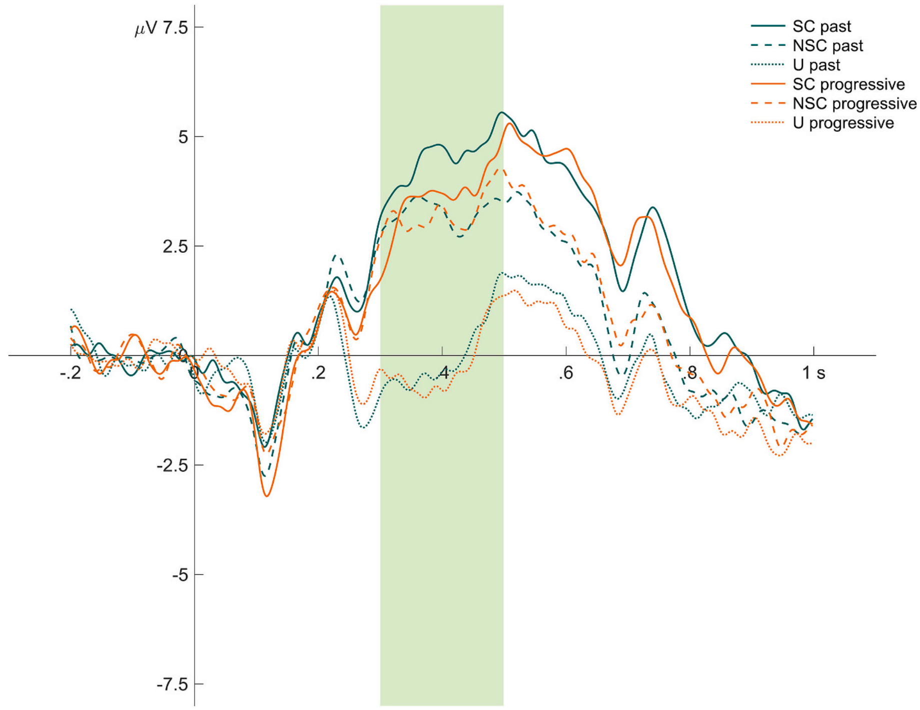

Fig. 3. Grand average waveforms collapsed across all six electrodes (Cz, CP1, CP2, P3/z/4) in the 300 to 500 millisecond time-window (highlighted).

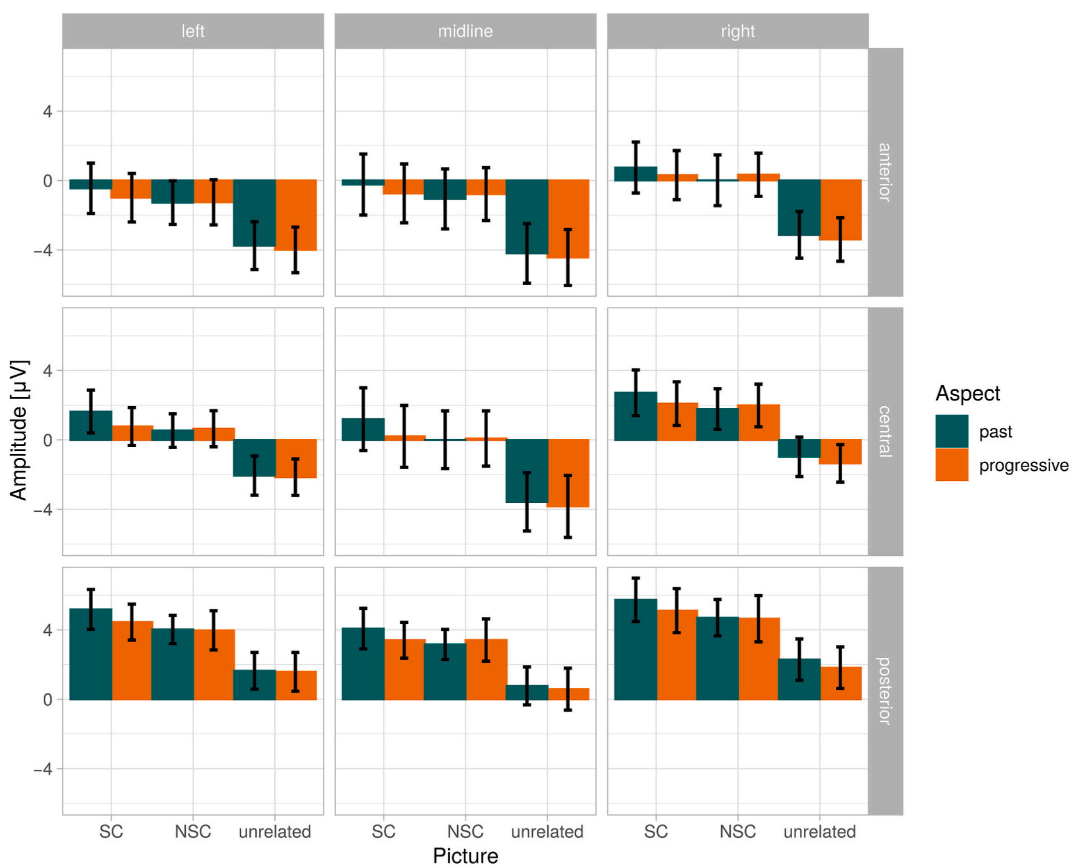

Fig. 4. Bar charts of the grand average amplitudes in the 300 to 500 millisecond time-window for the electrodes grouped according to the topographical factors Anteriority (anterior, central, posterior) and Laterality (left, midline, right) ± 1 SEM.

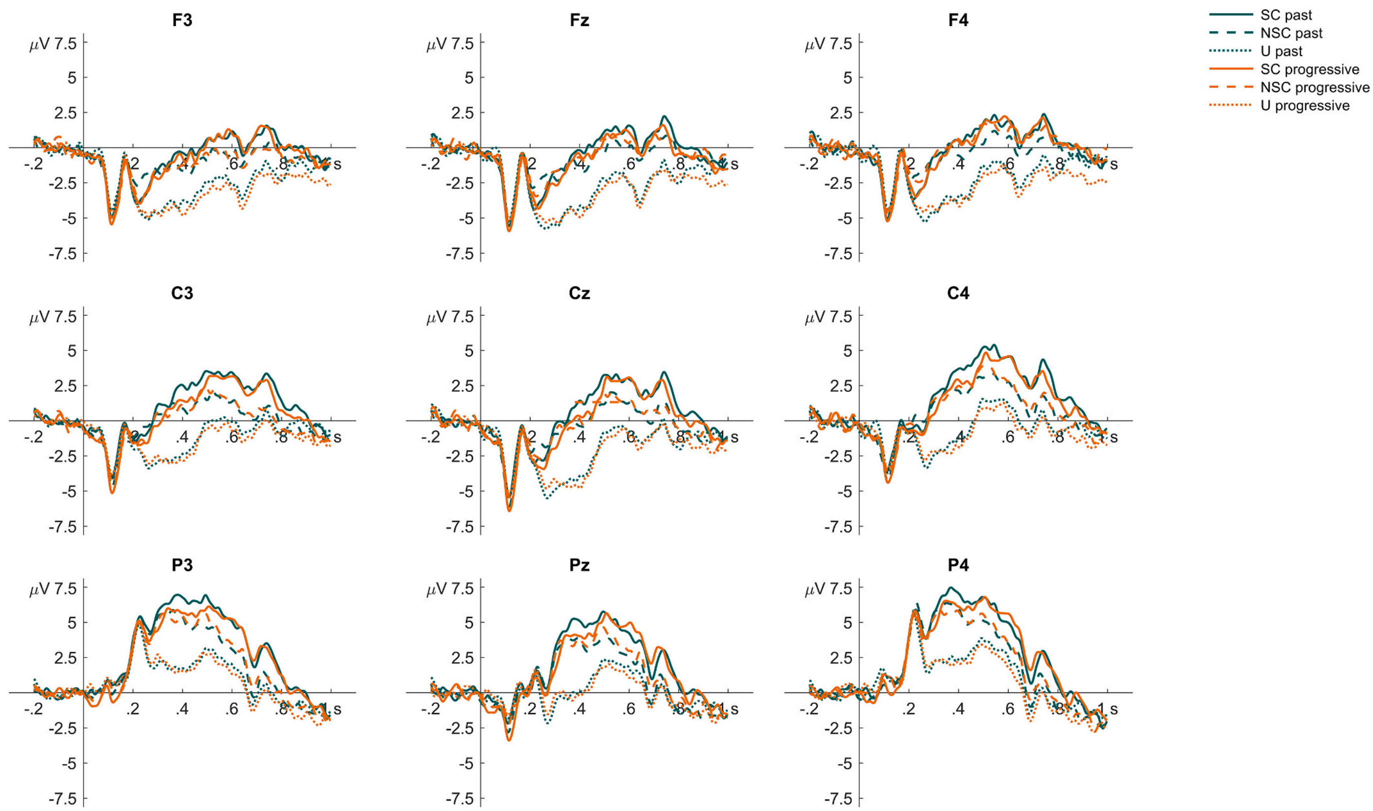

Fig. 5. Grand average waveforms at nine representative electrodes in frontal (F3/z/4), central (C3/z/4) and posterior (P3/z/4) areas.

2005) shows enhanced saliency of event and action goals and endpoints. The SC pictures in this study can be taken to represent the ultimate goal of the action (full object state change), and as such, this could have played a role in obtaining the observed main effect. Our findings are also in line with related work on the perception and memory of state-change events, showing enhanced visual attention to, and memory of, objects that undergo drastic change in events (e.g. Sakarias & Flecken, 2019; Santin, Van Hout, & Flecken, 2021), as compared to events that do not involve object state change (e.g., someone measuring a box, stirring in a bowl, that does not involve change in the physical properties of the objects involved). Note that the design meant participants saw the objects multiple times, so they may have been alerted to our state-change manipulation, and heightened an attentional focus on state-change. This in turn may have boosted the main effect of Picture. However, our main interest was in the interaction between Aspect and Picture type, which is further discussed below.

For the Aspect by Picture interaction, we observed differential processing for SC pictures after a sentence in the perfective as compared to a sentence in the progressive: SC pictures were perceived as more compatible with the sentence after the perfective as reflected by a high-amplitude P300, compared to the progressive. There was thus a lower degree of perceived match (Kok, 2001) between progressive sentences and SC objects. No difference in P300 amplitude for NSC pictures was found as a result of the Aspect manipulation. In theory, a holistic event representation as cued by the perfective would include the beginning state as well as the resultant state of an object undergoing change (Hindy et al., 2012; Solomon et al., 2015). However, given the verb semantics, and because of a potential underlying cognitive bias towards action goals, we find the modulation of aspect marking on event comprehension to be predominantly relevant to people's representations of the endpoint of an event, and thus to the resultant state of the object involved (i.e., the SC pictures). To assess the role of aspect in event comprehension further, follow-up work could also include in-progress pictures to get a clearer understanding of the object states available during event comprehension, in particular in light on the representations evoked by a progressive event viewpoint, which highlights intermediate stages of an event rather than its beginning and end states. To sum up, our findings regarding the differential processing of SC pictures are compatible with previous behavioural research in that they show that grammatical aspect matters for how we conceptualise sentence content and build event models (e.g., Madden & Zwaan, 2003, Madden-Lombardi et al., 2007). We have extended these findings by showing the implications of aspect marking for object representations in descriptions of change of state events. Further, our analyses of ERPs allowed a window on the cognitive processes relevant to sentence picture matching, prior to decision making.

According to Zacks (2019), language −here: aspect marking at the verb− may have an effect on how working models of events are constructed or how they are employed through different attentional mechanisms to guide expectations about how an event will unfold. Our results imply that participants' attention to the pictures reflected prior expectations of a *specific* visual instantiation of an object given what the sentence described. This is in line with Kang et al. (2019), who show that specific object states can be activated through the verb. According to IOH, objects and their state-changes are central to how we understand events. Specifically, Altmann and Ekves (2019) argue that relational binding, which enables us to track object state change, is mediated by attentional processes. Overall, the absence of the N400 for NSC pictures following sentences in the perfective suggests that participants may have activated both the multiple state representations of the object in this condition, which is in line with the activation of and competition between multiple object states during event comprehension. (Hindy et al., 2012; Solomon et al., 2015). In other words, NSC pictures were not evaluated as a general mismatch after encountering a description of a completed state-change event. Interestingly, as mentioned above there was no modulation of Aspect for the processing of NSC pictures. Note that NSC pictures clearly show the beginning phase of the described event. However, the verb semantics in the sentences despite highlighting different phases of the event cued by the Aspect, overall describe an action with an inherent endpoint (telic predicates). In other words, both sentences may have activated a number of object state representations, but in both cases the initial state was only weakly represented not leading to a competitive preference. By contrast, for SC pictures we do observe an effect of Aspect on the competition between object state representations: The perfective aspect at the verb constrained possible event representations leading to a specific attentional focus –and the SC picture "winning" the competition of possible representations–, whereas the progressive does not favour any of the two object representations tested here. This makes sense, given that the event is in progress, and the resultant state has not yet been achieved, nor is it clear where exactly we are within the process of inflicting change. Thus, SC pictures, which depict a specific event model (i.e., a completed change-of-state event) lead to differential responses. In sum, the specific verb form mattered for the degree to which both representations were active in participants' minds.

Events can be understood directly through experiencing them in the real world, and indirectly through language, for example in narrative texts. IOH is specific in that it discusses events and our understanding of them as bound to objects and their histories. In that sense, the state changes an object undergoes are essential to understanding an event. In our study, participants experienced the event indirectly *through* language with grammatical aspect manipulated at the verb level. The action described by the verb cued the state-change of the object while aspect situated the event (cf. Topic Time; Klein, 1994, 2004) at different points on a timeline, thereby focusing (perfective) or defocusing (progressive) state change, and rendering a mental image of a given degree of change. Experiencing the chopping of an onion through language, thus, arguably leads to different conceptualizations of an event as compared to experiencing the chopping of an onion directly. Though event cognition is an emerging field of study, the role of language in human ability to cognize about events is currently understudied. Our findings open up further possibilities for exploring the potential of language in cueing specific situation models during event perception, and also in event memory.

In conclusion, our study suggests that aspectual viewpoint expression can affect how change of state events are understood as they unfold, showing a subtle modulation of language concerning event comprehension.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

We thank Valeria Pfeifer and Jeroen Geerts for their assistance with the stimuli creation, Michel-Pierre Jansen and Peter Withers for their help with the implementation of the pre-tests, Maarten van den Heuvel for his support setting up the study, and Phillip Alday for his insights and help regarding the analysis using linear mixed effects models. We also thank the reviewers of the manuscript for their insights and helpful comments.

Appendix A. Appendix

List of verbs and the objects they were combined with in the SC and NSC conditions (objects in the U condition were drawn randomly for each participant from a pool of 84 objects, see table below).

Past progressive	Simple past	Object	Pool of unrelated object pictures
was beating	beat	the cream.	lego tower
		the butter.	flower
was biting	bit	into the sandwich.	scarf
		into the cookie.	cup
was boiling	boiled	the pasta.	paper plane
		the water.	nacklace
was burning	burned	the toast.	nail
		a match.	coca cola
was carving	carved	the pumpkin.	schnitzel
		wood.	bubblegum
was chopping	chopped	the onion.	sentence
		the leek.	shoes
was cracking	cracked	the chocolate bunny.	box
		the walnuts.	poem
was crushing	crushed	the pepper.	brownies
		the ice cubes.	bin
was cubing	cubed	the cucumber.	lollipop
		the cheese.	package
was dicing	diced	the tomatoes.	doll
		the garlic.	barbie
was dipping	dipped	the crackers.	email
		the bread.	book
was dissolving	dissolved	the powder.	backpack
		the pill.	puppet
was frying	fried	the mushrooms.	frisbee
		the steak.	puzzle
was grating	grated	the parmesan.	chips
		the carrot.	cushion
was grinding	ground	chillies.	socks
		coffee.	tennis ball
was halving	halved	the avocado.	dough
		the melon.	pencil
was juicing	juiced	grapefruits.	coins
		oranges.	clothes

(*continued on next page*)

(*continued*)

Past progressive	Simple past	Object	Pool of unrelated object pictures
was mashing	mashed	the peas.	keys
		the beans.	fork
was melting	melted	the ice cream.	shopping list
		the chocolate.	soup
was mincing	minced	parsley.	pear
		almonds.	cactus
was peeling	peeled	the banana.	ring
		the mandarin.	magazine
was roasting	roasted	the chicken.	card house
		the peppers.	house
was skinning	skinned	the potato.	hat
		the apple.	pot
was slicing	sliced	the bagel.	napkin
		the pizza.	bracelt
was squeezing	squeezed	the limes.	dowl
		the lemons.	milk
was breaking	broke	the vase.	waffle
		the clock.	candy
was coloring	colored	the drawing.	essay
		the painting.	knot
was destroying	destroyed	the files.	lamp
		the notes.	postcard
was lighting	lit	the cigarette.	hummus
		the candle.	lasagna
was scratching	scratched	the screen.	bucket
		the phone.	bonbon
was spilling	spilled	the paint.	plastic bag
		the whiskey.	teddy bear
was staining	stained	the bib.	balloon
		the blanket.	text message
was tearing	tore	the photo.	newspaper
		the shirt.	handbag
was shredding	shredded	the contract.	dice
		the bank statements.	rock
			crossword
			pudding
			skirt
			tights
			boomerang
			playdough
			crayon
			marbles
			toys
			wallet
			spoon
			salad
			smoothie
			plant
			juice
			cherries
			catalogue

References

Alday, P. M. (2019). How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *arXiv* (preprint arXiv:1707.08152).

Alday, P. M., & Kretzschmar, F. (2019). Speed-accuracy tradeoffs in brain and behavior: Testing the independence of P300 and N400 related processes in behavioral responses to sentence categorization. *Frontiers in Human Neuroscience, 13*, 1–22. https://doi.org/10.3389/fnhum.2019.00285.

Altmann, G. T., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*. https://doi.org/10.1037/rev0000154. Advance online publication.

Amoruso, L., Gelormini, C., Aboitiz, F. A., González, A., Manes, F., Cardona, J., & Ibanez. (2013). N400 ERPs for actions: Building meaning in context. *Frontiers in Human Neuroscience, 7*, 57. https://doi.org/10.3389/fnhum.2013.00057.

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science, 37*(2), 286–309. https://doi.org/10.1111/cogs.12006.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Becker, R. B., Ferretti, T. R., & Madden-Lombardi, C. J. (2013). Grammatical aspect, lexical aspect, and event duration constrain the availability of events in narratives. Cognition, 129(2), 212-220. Doi: https://doi.org/10.1016/j.cognition.2013.06.014.

Connell, L. (2007). Representing object colour in language comprehension. *Cognition, 102*(3), 476–485. https://doi.org/10.1016/j.cognition.2006.02.009.

Feller, D. P., Eerland, A., Ferretti, T. R., & Magliano, J. P. (2019). Aspect and narrative event segmentation. *Collabra: Psychology, 5*(1).

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33* (1), 182–196. https://doi.org/10.1037/0278-7393.33.1.182.

Flecken, M., & Gerwien, J. (2013). Grammatical aspect influences event duration estimations: Evidence from Dutch. In *, 35(35). Proceedings of the Annual Meeting of the Cognitive Science Society*.

Hindy, N. C., Altmann, G. T., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object state-changes on event processing: Do objects compete with themselves? *Journal of Neuroscience, 32*(17), 5795–5803. https://doi.org/10.1523/JNEUROSCI.6294-11.2012.

Hirschfeld, G., Feldker, K., & Zwitserlood, P. (2012). Listening to "flying ducks": Individual differences in sentence–picture verification investigated with ERPs. Psychophysiology, 49(3), 312-321. Doi; https://doi.org/10.1111/j.1469-8986.2011.01315.x.

Hirschfeld, G., Zwitserlood, P., & Dobel, C. (2011). Effects of language comprehension on visual processing–MEG dissociates early perceptual and late N400 effects. *Brain and Language, 116*(2), 91–96. https://doi.org/10.1016/j.bandl.2010.07.002.

Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., & Altmann, G. T. (2019). The influence of state change on object representations in language comprehension. *Memory & Cognition*, 1–10. https://doi.org/10.3758/s13421-019-00977-7.

Kelly, S. P., & O'Connell, R. G. (2015). The neural processes underlying perceptual decision making in humans: Recent progress and future directions. *Journal of Physiology, Paris, 109*(1–3), 27–37. https://doi.org/10.1016/j.jphysparis.2014.08.003.

Klein, W. (1994). *Time in language*. London: Routledge.

Klein, W. (2004). How time is encoded. In W. Klein, & P. Li (Eds.), *The expression of time* (pp. 39–83). Berlin: Mouton de Gruyter.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*(3), 557–577. https://doi.org/10.1017/S0048577201990559.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205. https://doi.org/10.1126/science.7350657.

Kuznetsova A., Brockhoff P.B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software*,* 82(13), 1–26. Doi: 10.18637/jss.v082.i13.

Lakusta, L., & Carey, S. (2015). Twelve-month-old infants' encoding of goal and source paths in agentive and non-agentive motion events. *Language Learning and Development, 11*(2), 152–175. https://doi.org/10.1080/15475441.2014.896168.

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition, 96*(1), 1–33. https://doi.org/10.1016/j.cognition.2004.03.009.

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

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition, 31*(5), 663–672. https://doi.org/10.3758/BF03196106.

Madden-Lombardi, C., Dominey, P. F., & Ventre-Dominey, J. (2017). Grammatical verb aspect and event roles in sentence processing. *PLoS One, 12*(12), Article e0189919. https://doi.org/10.1371/journal.pone.0189919.

Magliano, J. P., & Schleich, M. C. (2000). Verb aspect and situation models. *Discourse Processes, 29*(2), 83–112. https://doi.org/10.1207/S15326950dp2902_1.

Monroy, C. D., Gerson, S. A., & Hunnius, S. (2017). Toddlers' action prediction: Statistical learning of continuous action sequences. *Journal of Experimental Child Psychology, 157*, 14–28. https://doi.org/10.1016/j.jecp.2016.12.004.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 9. https://doi.org/10.1155/2011/156869.

Ostarek, M., Joosen, D., Ishag, A., De Nijs, M., & Huettig, F. (2019). Are visual processes causally involved in "perceptual simulation" effects in the sentence-picture verification task? *Cognition, 182*, 84–94. https://doi.org/10.1016/j.cognition.2018.08.017.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition*. Oxford: Oxford University Press.

Richmond, L. L., & Zacks, J. M. (2017). Constructing experience: Event models from perception to action. *Trends in Cognitive Sciences, 21*(12), 962–980. https://doi.org/10.1016/j.tics.2017.08.005.

Sakarias, M., & Flecken, M. (2019). Keeping the result in sight and mind: General cognitive principles and language-specific influences in the perception and memory of resultative events. *Cognitive Science, 43*(1), Article e12708. https://doi.org/10.1111/cogs.12708.

Santin, M., Van Hout, A., & Flecken, M. (2021). *Event endings in memory and language* (pp. 1–24). Cognition and Neuroscience: Language. https://doi.org/10.1080/23273798.2020.1868542.

Solomon, S. H., Hindy, N. C., Altmann, G. T., & Thompson-Schill, S. L. (2015). Competition between mutually exclusive object states in event comprehension. *Journal of Cognitive Neuroscience, 27*(12), 2324–2338. https://doi.org/10.1162/jocn_a_00866.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2), 153–156. https://doi.org/10.1111/1467-9280.00326.

Withers, P. (2017). *Frinex: Framework for interactive experiments*. https://doi.org/10.5281/zenodo.3522911.

Zacks, J. M. (2019). Event perception and memory. Doi: 10.31234/osf.io/634dz.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3.

Zwaan, R. A. (2008). Time in language, situation models, and mental simulations. *Language Learning, 58*, 13–26. https://doi.org/10.1111/j.1467-9922.2008.00458.x.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS One, 7*(12), Article e51382. https://doi.org/10.1371/journal.pone.0051382.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430.