Funding Body
This research was funded by the Max Planck Society for the Advancement of Science (www.mpg.de/en).

International Max Planck Research School (IMPRS) for Language Sciences
The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond. More information can be found at www.mpi.nl/imprs

The MPI series in Psycholinguistics
Initiated in 1997, the MPI series in Psycholinguistics contains doctoral theses produced at the Max Planck Institute for Psycholinguistics. Since 2013, it includes theses produced by members of the IMPRS for Language Sciences. The current listing is available at www.mpi.nl/mpi-series

ISBN: 978-94-92910-37-0
Cover design and layout: Julia Misersky
Print: Ipskamp

 The research reported in this thesis was conducted at the Max Planck Institute for Psycholinguistics, in Nijmegen, the Netherlands

About Time:

Exploring the Role of Grammatical Aspect in Event Cognition

Proefschrift
ter verkrijging van de graad van doctor aan de Radboud Universiteit Nijmegen op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken, volgens besluit van het college voor promoties in het openbaar te verdedigen op

woensdag 29 juni 2022
om 13.30 uur precies

door

Julia Misersky
geboren op 21 december 1987
te Berlijn, Duitsland

Promotor:

prof. dr. P. Hagoort

Copromotor:

dr. M.E.P. Flecken (Universiteit van Amsterdam)

Manuscriptcommissie:

prof. dr. C.F. Rowland

prof. dr. C. von Stutterheim (Universität Heidelberg, Duitsland)

dr. C. Madden-Lombardi (Université de Bourgogne, Frankrijk)

About Time:

Exploring the Role of Grammatical Aspect in Event Cognition

Doctoral thesis

to obtain the degree of doctor from Radboud University Nijmegen

on the authority of the Rector Magnificus Prof. Dr. J.H.J.M. van Krieken, according to the decision of the Doctorate Board to be defended in public on

Wednesday, June 29, 2022

at 13:30

by

Julia Misersky

born on December 21, 1987

in Berlin, Germany

PhD supervisor:

Prof. Dr. P. Hagoort

PhD co-supervisor:

Dr. M.E.P. Flecken (Universiteit van Amsterdam, The Netherlands)

Manuscript Committee:

Prof. Dr. C.F. Rowland

Prof. Dr. C. von Stutterheim (Universität Heidelberg, Germany)

Dr. C. Madden-Lombardi (Université de Bourgogne, France)

Contents

CHAPTER 1

GENERAL INTRODUCTION

"Events are a fundamental aspect of cognition. They give our thoughts and action purpose and are the basis of our intelligent understanding of the world."
—Radvanksy and Zacks (2014, p. 39)

This thesis aims to integrate event cognition theory with psycholinguistic research. First, I will discuss theoretical approaches to event cognition, and theoretical accounts of language comprehension, specifically on understanding linguistically described events. Then, I will discuss how time is encoded in language, and focus in more detail on grammatical aspect and the theoretical ramifications aspect may have for event cognition. Lastly, I will outline the thesis focus, namely how and under what conditions language—using grammatical aspect as a test case—guides event perception and memory.

1. Event Cognition

Over the last two decades, event cognition has become a research field of study in its own right. It explores how we understand the everyday happenings around us at the interface of perception, action, language, and memory.

How do we make sense of our daily experiences? Event Segmentation Theory (EST) proposes that we use predictive mechanisms to chunk our experiences into discrete events (Zacks & Swallow, 2007; Zacks & Tversky, 2001). With each new experience, a working model of the ongoing event is created, taking into account prior events of similar nature. Once it becomes too difficult to predict how the ongoing experience is going to unfold, we perceive an event boundary. This usually happens when a change occurs across one (or more) event dimension(s). The framing dimensions of any event are time and space. Additional dimensions, including entities such as people or objects involved in the event, further characterise the event. Objects specifically have been discussed as particularly central to event cognition (Radvansky & Zacks, 2010; Altmann & Ekves, 2019). A drastic change in space or time, or concerning entities such as objects, can lead to an increased inability to predict what is going to happen next. An event boundary is detected—i.e., the event has ceased—, and anything following it will be understood as the beginning of a new, distinct event. Generally speaking, then, each event ends up as a segment of time with beginning and end. The working model becomes an event model, a mental representation of the event in memory, ready to be employed again and again when we make new experiences. Note that this is a simplified way of describing event segmentation, as event dimensions must be considered in relation to one another, and change tends to happen simultaneously across domains. This is illustrated below by examples of two distinct event types, which will be further explored in this thesis:

> (1) Imagine an onion being chopped. The changes the onion undergoes inevitably are in relation to time passing, but also in relation to the properties and affordances of the onion itself as well as the agent doing the chopping.
>
> (2) A person walking across a field towards a house: The agent's movement on the trajectory across the path (the field) happens over time, with the agent's location and their relation to the goal (the house) continually changing.

As outlined by the above examples, change always occurs at more than one dimension, meaning different dimensions jointly contribute to the detection of event boundaries. Testing this empirically, Huff, Meitz and Papenmeier (2014) had participants watch sitcom scenes with changes across one, two, three, or four dimensions. Participants had to segment the scenes into meaningful units (event segmentation task; Newtson, 1973). The authors also tested memory for the sitcoms. Results indicate that while situations closer to an event boundary are remembered better the more changes are perceived, an increase in the number of changes also makes predicting what is going to happen next more difficult.

Importantly, theoretical accounts of event cognition acknowledge language as a means to experience, understand and remember events. Radvansky and Zacks (2014) discuss the term 'situation model' as an event model derived from language, which during language comprehension may serve as a mental simulation of an event and allow for predictions about how it may unfold. Language can express change in events in a number of different ways and linguistic descriptions are found everywhere from listening to the news to reading a novel to an email sent to a friend. We read, hear, and talk about our day-to-day experiences often.

This thesis aims to explore (1) to what extent language can activate specific visual representations descriptive of an event, (2) whether such representations can be activated from limited language contexts, (3) how language can affect memory for events, and (4) to what extent previous work on linguistic experience and event cognition generalises to ecologically valid settings. In short, whether and to what extent language shapes our event models and with it our understanding of events, will be the core hypothesis of the research outlined in this thesis.

2. Event Cognition and Language

The interplay of event cognition and language can take two directions: Firstly, previously experienced events can guide language comprehension, because they can serve as cues as to how an event expressed through language may unfold (e.g., Altmann & Kamide, 1999; Altmann & Mirković, 2009). Secondly, and most relevant to this thesis, language can also guide event understanding by providing cues for event model construction. As will be discussed below, these two directions are strongly related as they both draw on similar predictive processes underlying model construction for comprehension.

In their discussion on situation models in language comprehension, Zwaan and Radvansky (1998) outline five dimensions which characterise situations as expressed in language (time, space, intentionality, causation, and protagonist). Similar to event models derived from experience then, language-based situation models make use of spatio-temporal information (cf., Zacks & Tversky, 2001) to guide successful language comprehension by providing a framework in which incoming language input can be integrated. Any linguistic description of an event usually contains a verb, and according to Zwaan, Langston and Graesser's (1995) Event-Indexing Model any event in language (as denoted by a verb) is composed of the aforementioned five dimensions. During language comprehension, newly incoming linguistic information is checked for similarity along these dimensions; higher similarity across dimensions means the new input is more easily integrated with the situation model. Simplified, then we can say that when a sentence describes a situation at a particular (event) time X, and a new sentence does not suggest any changes in time X, both sentences are likely to be integrated within the same situation model. Note here that the dimensions are related and may jointly contribute how a situation is understood, which is why Zwaan and Radvansky (1998) argue that dimensions should not be studied in isolation. In line with this, empirical evidence suggests that the more dimensional changes are signalled by linguistic input, the longer people take to process said input, which is reflected in slower reading times (e.g., Curiel & Radvanksy, 2014). Constructing a situation model thus becomes more effortful the more changes occur at relevant event dimensions, which is in line with the abovementioned results by Huff et al. (2014). Noticing the change described by language is tightly coupled to the linguistic means with which the content is conveyed (e.g., grammatical marking, verb semantics etc.), but also to our knowledge of similar events (Zwaan & Radvansky, 1998). Using an event segmentation task, in which people had to flag breakpoints (i.e., when one unit ends and another begins) in an event, Zacks, Tversky and Iyer (2001) have shown that people's perception of events reflects a hierarchical organisation of fine (e.g., to sieve flour, weigh out sugar) and coarse units (e.g., to bake a cake). This hierarchical structure is also reflected, and even strengthened, by how people describe event units and how familiar they are with events. In this sense, language seems to reflect and interact with the perceived event structures, again suggesting that model construction in event cognition and language comprehension appear to be based on similar cognitive mechanisms.

Some psycholinguistic work specifically examined the interplay of event cognition and linguistic experience. For example, Gerwien and von Stutterheim (2018) presented native French and German speaking participants with motion video clips. In Experiment 1, participants had to describe the clips. The data revealed that German speakers were more likely to describe the events in one event unit (one clause) focusing on direction (e.g., "*Eine Frau geht um einen Brunnen herum eine Treppe rauf*", '*A woman goes around a fountain hither-round a 'set of steps' up.*') compared to French speakers. This is because such a change in direction cannot be encoded in French in a single clause, with the consequence that French speakers need to use more clausal units to describe the same event. The results thus reflect grammatical differences regarding how motion is encoded at the level of the verb. Interestingly, these differences also show up in how participants segment events; French speakers had a higher instance of flagging an event boundary (i.e., they segmented more frequently) compared to German speakers. The event units used to describe motion direction mapped onto the perceived event units as reflected in segmentation patterns across the two language groups.

While the above examples describe how language is related to event boundaries, some work, which will be reviewed in more detail throughout the chapters of this thesis, also highlights that language can affect understanding of event content. For example, Sauppe and Flecken (2019), show that sentence structure (active vs passive sentences) guides eye gaze to specific event features during first apprehension. Further, specific descriptions can affect how people judge the similarity of events (e.g., Montero-Melis, Jaeger & Bylund, 2016; Athanasopoulus & Bylund, 2013), and language use can affect how events are memorised (e.g., Sakarias & Flecken, 2019; Skordos et al., 2020). Generally speaking, observed effects of language on event cognition appear to be stronger in verbal compared to non-verbal tasks.

Importantly, though, the above emphasises that theories of event perception and language comprehension show remarkable similarities when it comes to the structure and dimensions of event (or situation) models[1] (for a review, see Ünal, Yi & Papafragou, 2019). Understanding an event's structure is paramount to understanding the event, whether experienced directly or indirectly. As such, 'Language provides unique opportunities to study event comprehension more broadly [...]' (Radvanksy & Zacks, 2014, p. 79). However, despite the abovementioned similarities, only few psycholinguistic studies explicitly and systematically take event cognition theory into account.

3. Event Time in Language

Returning to the framing pillar of event time, language allows us to 'place events in time [...]' (Radvanksy & Zacks, 2014, p.72). Reflecting characteristics of Event Segmentation Theory, Klein's notion of time in language specifies that time can be segmented into smaller chunks (Klein, 2009). In line with this, Madden and Ferretti (2009) highlight that any description of a situation inevitably carries information about time. There are a number of ways in which time can

[1]The terms situation and event model will be used interchangeably throughout this thesis, unless a specific theoretical deliberation requires detailed discussion of their distinctive meanings.

be conveyed at the level of the verb[2], highlighting the temporal features inherent to the situation at hand. For one, verb semantics generally denote situations, which may differ regarding their temporal properties. This allows for categorisation of verbs into different verb classes (called the category of *Aktionsart*, or lexical aspect; see Madden & Ferretti, 2009 for an overview). We can distinguish the following four verb classes: states, events, processes, and accomplishments (which have properties of both processes and events). Verb classes differ on a number of semantic dimensions, including, among others, whether they denote situations that exist or that happen, and whether they refer to situations with (telic) or without (atelic) inherent endpoint (Comrie, 1976). The arguments of the verb and other elements within the verb predicate (e.g., verb particles) can render distinct temporal interpretations of a situation, compare e.g., "*to chop an onion*" (telic situation, with an inherent endpoint) to "*to chop onions*" (atelic event, without a specified endpoint) and "*to chop off the branches of a tree*" (telic event). Depending on the language and verb under discussion, however, verb classes may overlap in their semantic dimensions making universal agreement of the categorisation of situation types somewhat difficult. Note that throughout this thesis, the term 'event' is used in line with event cognition theory to refer to a situation in the outside world, rather than in terms of verb class.

Additionally to *Aktionsart* and the associated verb classes, time can be encoded in language through *tense*, a means to situate a description in relation to the utterance (usually, the now), and—most relevant to this thesis—*aspect*, allowing for a specific viewpoint on a situation. Note that tense and aspect function independently from another, which will be outlined below. Radvanksy and Zacks (2014) acknowledge the role of grammatical aspect to guide event model construction and event segmentation, as aspect relates a situation to the framing pillar of time. Most commonly, we can distinguish between perfect, perfective, and imperfective (or progressive) aspect, characterising an event with focus on the resultant state, or as bounded or as ongoing, respectively. Both perfect and perfective aspect make reference to event completion. In line with previous empirical work on grammatical aspect, the focus of this thesis will be on the contrast between perfective and progressive aspect, with the majority of the thesis focusing on English.

[2]There are additional ways time can be conveyed by language, including adverbials (e.g., "*soon*", "*yesterday*" etc.) or—depending on the language at hand—particles and discourse organisation, though these means are not discussed here in further detail.

Relating this to Klein's notion of time, we can exemplify this distinction between perfective and progressive below (Fig. 1):

Figure 1: *Exemplification of Klein's notion of Topic Time, and how perfective (top) and progressive (bottom) aspect relates Topic Time to an event with an inherent duration (e.g., to chop the onion; TSit).*

The above highlights two sentences in the past tense. The past locates the 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. As can be seen in Fig. 1, the situation or event has an inherent duration (TSit). Topic Time refers to 'the time for which the particular utterance makes an assertion' (Klein, 1994, p.37). Aspect, in English morphosyntactically realised at the verb, thus relates TSit to Topic Time. It can specify Topic Time and accordingly define the phase under discussion; in the case of the perfective, the event is represented as bounded, and the endpoint or completion stage is fully included in Topic Time. This is not the case for the progressive. The progressive can specify the internal phase(s) of Topic Time but not necessarily make reference to the (beginning and) endpoint[3]. Instead, it 'stretches' the time span of events, thereby honing in on the intermediate phases of an (ongoing) event (Hong, Ferretti, Craven, & Hepburn, 2019; Madden-Lombardi, Dominey, & Ventre-Dominey, 2017; Matlock, 2011). Note also that languages may differ in their expression of ongoingness. In aspectual languages, such as English, aspect is grammaticalised at the verb. The distinction between completion and ongoingness may be particularly explicit for situations in which progression can be 'measured' (Flecken, 2010), as is the case for events characterised by degree of goal orientation. Returning to the previous examples, we can say the following:

---

[3]Note that while aspect can make reference to a situation as completed or ongoing, the effects of aspect may differ as a result of tense (cf. Madden & Ferretti, 2009 for further discussion), temporal information in the context such as adverbials (e.g., Flecken, Walbert, & Dijkstra, 2015 for empirical evidence on aspect mismatches), verb class, and due to differences across languages.

(1) In an event such as 'to chop an onion', the event starts out with an onion in its original state, and the goal is to transform the onion into a chopped version consisting of smaller pieces. In order to achieve this goal, a chopping action needs to be carried out, which irreversibly changes the onion until the goal state is reached. Throughout the chopping event, the (visual) changes the onion goes through can thus be taken as measures of progress towards the goal state.

(2) In an event, where a person is walking across a field towards a house, progress can be measured by following the agent's movement on the trajectory across the path (the field). The agent's changing location and their relation to the goal (the house) specifies the progress.

In both examples, the possibility to pinpoint and keep track of the progress allows for the distinction between ongoingness and completion of the event. By contrast, an event such as 'to sit in a chair' has no inherent progress as it describes a state, with the result that the distinction between ongoing and complete is less distinct than in the above examples.

4. Event Models and Aspect

Klein's theoretical notion of time in language links well to event cognition theory, and specifically Event Segmentation Theory, with recent empirical research showing that aspect marking can affect event segmentation. In their Experiment 1, Feller, Eerland, Ferretti and Magliano (2019) presented native speakers of English with stories including sentences in the perfective or progressive. Participants had to indicate for each sentence whether it signalled a change in the events of the stories. The results indicated that participants were more likely to flag sentences in the perfective as signalling a boundary compared to sentences in the progressive. As noted earlier, sources of temporal information can interact, and Experiment 2 thus used a similar setup, but included events that differ in their duration and can be classified as short (e.g., writing a check) or long (e.g., watching a movie; event duration measures based on Magliano & Schleich, 2000, who show that aspect affects perception of the duration of events). The results of Experiment 1 were replicated, with inherent event duration showing no effect. However, effects of inherent duration were observed in Experiment 3. In a continuation task, participants were asked to continue a short story consisting of a few leading sentences. When provided with leading sentences in the progressive, participants were more likely to write continuations that would fall during the event described in the lead sentences. Contrary, when the lead sentences were in the perfective, participants were more likely to write continuations describing what followed after the event described by the leading sentences. Effects of aspect on event duration have also been found in Dutch speakers (Gerwien & Flecken, 2013). Although in Dutch aspect is not grammaticalised, it can be realised lexically. When presented with the progressive, Dutch speakers tended to judge (inherently short) events as longer than when they were presented in the simple present tense; in contrast, inherently long events were estimated as lasting shorter in the progressive, as compared to the simple present.

Recall that situation models are event models derived from linguistic input, allowing us to predict how an event may continue. Radvansky and Zacks (2014) state that unless there is a specific event break signalled through the (linguistic) input, i.e., a boundary is detected, people have the tendency to integrate any additional information into the event at hand. In addition, Radvansky and Zacks (2014) note that language may lead to people inferring information about event entities such as characters or objects. Aspect may play an important role with regards to which information is inferred, as empirical research has shown that information about event characters and event location is more available when linguistic content conveyed information in the progressive compared to the perfective (Carreiras, Carriedo, Alonso, & Fernandez, 1997; Ferretti, Kutas, & McRae, 2007).

Looking at in-progress versus ceased accomplishments, and using a sentence-picture verification task, Madden and Zwaan (2003) presented participants with sentences describing an event in the perfective (simple past; "*made a fire*") or progressive (past progressive; "*making a fire*"). In Experiment 1, the picture was followed by two picture alternatives, depicting the event as ongoing or completed. Participants were to choose which picture matched the sentence. Their choices showed that participants were more likely to pick the completed picture after having read the perfective sentence compared to the ongoing picture. No such match preference was observed for sentences in the progressive. Experiment 2 showed a similar pattern of effects; here, only one picture choice was provided after the sentence. Reaction times revealed that after reading the perfective, participants were quicker to decide the completed picture was a match compared to the ongoing picture. There was no such match advantage for the progressive, which Madden and Zwaan (2003) interpret to be due to variability between the participants, suggesting that what they represent upon reading an event in the progressive is less specific compared to what is represented when a perfective event is read. Also using the sentence-picture verification task, Yap et al. (2009) tested the interaction between aspect and different event types. While Madden and Zwaan (2003 used accomplishments, i.e., events with an inherent endpoint, Yap et al. (2009) specifically contrasted accomplishments (e.g., knitting a sweater) and activities (e.g., swimming). They presented Cantonese speaking participants with sentences in the perfective or progressive, followed by a picture pair. The task was to indicate which sentence they just heard by choosing one of the pictures. Similar to Madden and Zwaan (2003), they found perfective facilitation with accomplishments. They expanded Madden and Zwaan's (2003) findings by observing progressive facilitation for activities.

Building on this, Becker, Ferretti and Madden-Lombardi (2013) presented English speaking participants with passages describing accomplishments (e.g., to pack a lunch) or activities (e.g., to exercise), in which a critical sentence was presented either in the perfective or the progressive. The critical sentence was followed by a sentence suggesting a long or short time had passed, after which the target word (e.g., the lunch) was mentioned again. Using electroencephalogram (EEG), Becker et al. (2013) assessed how the second mentioning of the target word was processed. The results for accomplishments revealed that the temporal information conveyed by aspect interacted with the duration: For short events, people had more difficulty integrating the target word after the perfective (i.e., a larger N400, a marker for semantic integration) compared to the progressive. There was no such interaction of aspect for long events, and furthermore no differences observed for activities. Similar to the findings by Feller et al. (2019), the results highlight that temporal information such as event duration or time passing can interact with the temporal marking associated with aspect. In addition, the findings point towards effects of aspect being particularly relevant for accomplishments, which do have a natural endpoint, over activities, which do not. This aligns well with the aforementioned notion that progression is internally 'measured'. In that sense, this 'measurement' of progress seems to be related to the endpoint and the associated change in the event, which is more pronounced for some events but less so for others.

In line with the focus on completion and ongoingness of the perfective and the progressive, respectively, some studies have examined the effects of aspect on memory. For example, Hart and Albarracín (2009) had participants engage in anagram solving tasks. In the experiments, they then disrupted the participants' ongoing task and prompted them to describe their behaviour during the anagrams either in the progressive or perfective. In Experiment 2, participants then were engaged in a recognition memory task for anagrams, having to correctly categorise previously seen ones from unseen ones. The authors found that memory accuracy was better when they described their behaviour using the progressive. Moreover, participants were also more willing to resume the anagram task, when they used the progressive compared to the perfective. These results are in line with findings suggesting detailed information about an event is more available in ongoing event contexts (e.g., Carreiras et al., 1997). Looking at event memory more specifically, Salomon-Amend, Radvansky, Anderson, Zwaan and Eerland (2017), recently conducted a web-based study, in which native speakers of English read sentences (e.g., "*The mother laughed/was laughing at the joke.*"). Participants' verbatim memory for the sentences was tested using three different memory tests: a Yes/No recognition task of the sentences (Experiment 1), a two-alternative forced choice (2AFC) task asking participants to indicate which sentences they had previously read (Experiment 2), and a cued recall task asking participants to reproduce the sentences (Experiment 3). In all three experiments, memory accuracy for the sentences was better after the perfective than the progressive. The authors discuss their results in terms of a cognitive bias to process events in forward order toward a completion point. As a result, memory is biased towards completed events, consistent with perfective aspect. This line of reasoning is supported by their results that progressive verbs are more likely to be misremembered by participants as having been perfective. At first sight, Salomon-Amend et al.'s (2017) findings seem at odds to Hart and Albarracín (2009). Note that in Hart and Albarracín's (2009) study, participants were engaged in a still ongoing event (i.e., the task at hand). By contrast, in Salomon-Amend et al.'s (2017) study, participants were constructing event models from sentences in the past. This interplay of real-world time, tense, and aspect may have led to the difference in result patterns. However, as Salomon-Amend et al. (2017) note, the work on aspect and memory, and specifically event memory, is sparse and requires additional targeted studies.

Overall, these examples highlight a general tendency for aspect to affect our perception of event timing and properties regarding event entities such as objects. This includes the duration of an event, but importantly also whether or not the event completion (or right boundary) is considered during event comprehension. In line with linguistic theory (Klein, 2009, 2014), empirical work attests that the perfective highlights event completion, whereas the progressive highlights internal event features.

5. Background of the Thesis Research

5. 1. Change-of-State Events and Aspect

In event cognition theory there has been a recent focus on object change as particularly relevant to understanding events. In contract to Event Segmentation Theory (Zacks & Swallow, 2007), the Intersecting Object Histories account (IOH, Altmann & Ekves, 2019) is less concerned with the detection of boundaries but focuses on how we understand what happens within these boundaries. According to IOH, we make sense of the internal event structure by tracking the change an object undergoes throughout an event. By tracking object states, we can also make reference and understand the changes across other event dimensions, such as the people within the event, or event time. Both EST and IOH are two sides of the same coin, both relying on the same predictive mechanisms but describing different facets of event cognition.

A neuroimaging study (Hindy, Altmann, Kalenik, & Thompson-Schill, 2012) looked at the tracking of objects states in two experiments. Specifically, using fMRI, the authors wanted to assess whether the changing object states throughout an event compete or not by manipulating degree of object change. In Experiment 1, they manipulated whether an object changed (e.g., "*chop an onion*") or not (e.g., "*weigh an onion*") as a result of an action. In Experiment 2, the action was held constant, but the affordance of the associated objects either reflected a change (e.g., "*stomp on an egg*") or not (e.g., "*stomp on a penny*"). Crucially, in an event comprehension task, participants were presented with one sentence that suggested a state change or not, and a second sentence that always suggested the object being minimally affected (e.g., "*The chef will chop/weigh the onion. And then, she will smell the onion.*"). Fillers were used as catch trials, to which participants were asked to respond via button press. Following the event comprehension task, Hindy and colleagues (2012) identified brain areas responsive to conflict using the Stroop task and found activated voxels in left posterior Ventrolateral Prefrontal Cortex (pVLPFC). Their results indicate that degree of state change predicted the BOLD response in this area. This suggests that participants experience conflict during event comprehension as multiple object states are activated at the same time and appear to compete. Solomon, Hindy, Altmann and Thompson-Schill (2015) built on Hindy et al.'s (2012) work using a comparable setup and were able to replicate these findings. They emphasise further, that competition between object states occurs for single object tokens. Building on this work on object state change, and based on Zwaan and Pecher (2012), Kang, Eerland, Joergensen, Zwaan and Altmann (2019) used a sentence-picture verification task to test how people understand change-of-state events. In Experiments 2 and 3, they presented participants with sentences in the past or future tense and manipulated verb semantics to suggest a state change ("*dropped/will drop the ice cream*") or no state change ("*chose/will choose the ice cream*"). Participants then saw a picture probe showing the object in its original or changed state, and had to indicate whether the object in the picture was mentioned in the sentence. In both experiments, participants responded faster to the changed picture (compared to the original state picture) when the verb suggested state change. In addition, for Experiment 2, such a match effect was also observed for original state pictures, indicating that participants were faster to respond to original state pictures when the verb suggested no change. In sum, the degree of change suggested by verb semantics affects how the event is represented through the activation of (multiple) object states.

Madden-Lombardi et al. (2017) presented French native speaking participants with sentences (perfect vs progressive), where the critical instrument (in-use vs not-in-use) as well as the target objects (ongoing vs completed) were replaced by images. In order to control for context effects due to sentence order (and hence position of instruments and objects), two experiments were run; in Experiment 1, instruments appeared first in the sentence and in Experiment 2, the order was reversed such that the object appeared first in the sentence. Participants read each sentence word-by-word (advancing sequentially by pressing the spacebar). At the end of the sentence, they would read "*Acceptable?*" and had to indicate a Yes/No choice via button press. In line with their research question, reaction time was expected to differ according to the manipulation, since sentences contained either an instrument or object picture that did not match the aspect of the verb. Yes responses for such mismatches were expected to lead to slower reaction times (RTs). Results of Experiment 1 indicate that participants were faster to respond to congruent items (i.e., when aspect and picture of instrument/object matched) compared to incongruent ones. Furthermore, perfect sentences led to a match effect strongly driven by objects as compared to instruments. In Experiment 2, participants again were fast to respond to congruent items overall. Again, the match effect for perfective sentences was strongest for objects. Match effects for the progressive sentences in both experiments were marginal yet showed a small advantage for instruments over objects. Overall, the match effect for objects in perfective sentences highlights that perfective aspect focuses strongly on the goal or endpoint of the event.

As previously highlighted, aspect can have an impact on whether events are understood as ongoing or completed, putting emphasis on different phases of the event. In line with this, aspect may affect which features of an event are foregrounded, affecting event entities such as event participants, location, and objects (for a review see Madden & Ferretti, 2009). Linking to event cognition theory, aspect seems to have an influence on event segmentation by specifying the degree of change along an event dimension, which is in line with experimental findings (Feller et al., 2019). Specifically, the progressive signals ongoingness, and internal event features tend to be more available to a comprehender. By contrast, these features are less available when the perfective draws the focus towards the event's endpoint. Arguably, internal event properties are less relevant to understanding a perfective situation, where change is anchored to a resultant endpoint of an event. Relating back to the idea of tracking object states, aspect may thus be particularly influential for the understanding of events in which an object undergoes substantial and irreversible state change. In short, it is plausible that grammatical aspect can play a role regarding which object states are highlighted. With the exception of Madden-Lombardi et al. (2017), this has not been explored empirically in great detail—a gap the studies in this thesis aim to bridge.

5. 2. Motion Events and Aspect

In contrast to the work on grammatical aspect and change-of-state events, there is already a substantial body of literature on the effects of aspect on motion events.

In a series of studies, Anderson and colleagues explored the role of aspect in native speakers of English. In their 2008 study (Anderson, Matlock, Fausey, & Spivey, 2008), they used a mouse-tracking design, where participants were presented with scenes showing a character, path, and a potential goal destination in the distance. Participants heard sentences, and were then asked to place the character onto the scene to match the description. When participants heard sentences with verbs in the progressive (e.g., "*was walking*"), they were more likely to drop the character on the path, whereas the character was dropped closer to the destination when participants heard sentences in the perfective (e.g., "*walked*"). Aspect also affected how long participants spent on moving the character: movement duration (from grab-click to drop-click) was longer after sentences in the progressive compared to the perfective. In additional work (Anderson, Matlock, & Spivey, 2010), the sentences also included terrain descriptions (easy vs difficult). The original effect concerning drop locations was replicated. However, terrain descriptions only affected movement duration for sentences in the progressive. Specifically, participants' movements were faster when the terrain was described as easy compared to difficult. In 2013, Anderson, Matlock and Spivey also presented work, in which the temporal context was manipulated as recent ("*yesterday*") or distant ("*last year*") in addition to grammatical aspect. Again, the effect of aspect on drop location was replicated. For movement duration, the authors observed a cross-over effect: Progressive aspect paired with recent context, as well as perfective aspect with distant context, both led to faster movements. Anderson et al. (2013) describe this in terms of a match effect, and conclude that the perfective and distant context both share the tendency to describe the event holistically (or bounded) with little emphasis on event details. By contrast, the progressive allows for event details, which are also more available in a recent temporal context. Overall, their work emphasises that aspect in English affects event processing.

Von Stutterheim, Andermann, Carroll, Flecken and Schmiedtová (2012) used a cross-linguistic design to further study the effects of aspect marking on event perception, specifically focusing on motion events. Following the above line of reasoning, progressive aspect can 'stretch out' event perception, but it is important to note that aspect marking is not grammaticalised (and hence not obligatorily used) across languages. Von Stutterheim and colleagues (2012) presented speakers of languages with aspect (e.g., English, Russian) and without aspect (e.g., German) with video clips in which agents approached but did not reach a potential goal (e.g., a person running towards the train station, but the video ending before the person running actually reaches it). Participants had to describe the events after the video had stopped playing, and their eye gaze was being tracked. Speakers of non-aspectual languages were more likely to mention a destination in their descriptions, and also looked more often and longer towards the goal, compared to speakers of aspectual languages. Although speakers of non-aspectual languages can express ongoingness, they are unable to formulate a complete description if they omit the (potential) destination. As a result, speakers of non-aspectual languages tend to mention goals more frequently. Importantly, the results by von Stutterheim et al. (2012), show a continuum: Within speakers of non-aspectual languages, Dutch speakers, for example, were less attentive to goals compared to German speakers. This may be due to Dutch speakers frequently expressing ongoingness through lexical means (*aan het* construction). Flecken, Carroll and von Stutterheim (2014) replicated the abovementioned findings by studying the two extremes on the previously observed aspect continuum: German as a non-aspectual language, and Modern Standard Arabic (MSA) as an aspectual language. Using the same motion video clips and again employing eye tracking, they found that German speakers attended more to potential goals or endpoints compared to MSA speakers, who instead focused more on the path. The focus on endpoints for speakers of non-aspectual language was also reflected in the descriptions of the events; German speakers mentioned endpoints more frequently than MSA speakers did.

Cross-linguistic differences in motion event descriptions have additionally been observed in a comparison of English (aspectual) versus Swedish (non-aspectual) speakers (Athanasopoulos & Bylund, 2013). Native speakers of Swedish mentioned potential endpoints more often than native speakers of English. This finding was also reflected in a similarity judgment task performed by a new set of participants. In this nonverbal task, participants watched triads of video clips (adapted from von Stutterheim et al., 2012). First, they saw clip A, then clip B, and finally clip X. Participants had to decide whether X was more similar to A or B. Clip X always showed an intermediate goal orientation reflecting a true midpoint, whereas the alternates (A or B) showed low and high goal orientation, respectively. Swedish speakers opted for the alternative with high goal orientation more frequently than English speakers. The ABX task was employed also in a cross-linguistic comparison of Afrikaans and Swedish, both non-aspectual languages, to English as an aspectual language. Both speakers of Afrikaans and Swedish showed a similarly strong preference for the high goal orientation alternative, and this preference was significantly different to that of English speakers. English speakers chose the high goal orientation alternative less often. Overall, there was more variability across the Afrikaans speakers, which the authors interpret to be due to their frequent use of English. This is in line with von Stutterheim et al. (2012), who showed a variable effect of aspect depending on how aspect is realised in a specific language (i.e., grammatically or lexically), and the frequency of use of (lexical) expressions of ongoingness.

Cross-linguistic differences concerning motion event perception have also been studied using electrophysiological measures. In an ERP study, Flecken, Athanasopoulos, Kuipers and Thierry (2015), presented native speakers of German (non-aspectual) and English (aspectual) with animations, which depicted basic motion events exemplified by a dot on a trajectory towards another shape, yet never reaching it. In Experiment 1, participants first saw the animation and then a target picture, for which they had to indicate whether it matched the animation. There were four conditions, in which target and animation were either a full match (5% of trials), a trajectory match, an endpoint match, or a full mismatch (75%). Looking at the P300, an ERP relevant to attentional processing (Polich, 2007; Kok, 2001), they found that German speakers showed a larger P300 response in the endpoint match condition compared to the trajectory match condition, suggesting heightened attention to endpoints during event perception.

In sum, speakers of non-aspectual languages are more likely to show a bias towards potential goals or endpoints when watching motion events, which has been observed in online and offline measures. By contrast, speakers of aspectual languages do not show such a bias, and instead attend towards features of a scene that signal ongoingness since the progressive 'illuminates the path, not the destination.' (Salomon-Amend et al., 2017, p. 2).

6. Thesis Outline

In Chapter 2, I aimed to answer the question to what extent language can activate specific visual representations descriptive of an event. Based on previous work on event and object state representation, I have adapted a sentence-picture verification task for EEG. The paradigm has been previously used to study how people simulate events, including how they visually represent event entities (Madden & Zwaan, 2003; Hirschfeld, Zwitserlood & Dobel, 2011; Zwaan & Pecher, 2012, Kang et al., 2019). Opting for this task allowed me to build on existing work while specifically addressing the role of aspect in event model construction of change-of-state events. Participants were presented with sentences following a SVO structure, in which grammatical aspect was manipulated (e.g., "*John chopped/was chopping the onion*."). After sentence presentation, they saw a picture of an object. The object either showed the object in its original, unchanged state or in its changed state, and participants had to indicate whether the object in the picture was mentioned in the previous sentence. ERPs were measured from picture onset. Unrelated object pictures were included to serve as a control. Using EEG as a measure allowed for the fine-grained study of online event processing prior to decision-making, specifically focusing on events where an object undergoes change.

Chapter 3 used a comparable setup with the same stimuli as Chapter 2, but recording behavioural measures (reaction times, RTs). Experiment 1 predominantly served as an extension of Chapter 2 and allowed me to investigate in more detail the role of grammatical aspect in the understanding of change-of-state events in the face of conscious decision-making. In Experiment 2, I studied whether visual representations descriptive of an event can be activated from limited language contexts. To this end, I opted to use a limited sentence context and omitted the object noun. With event cognition theory, and specifically IOH (Altmann & Ekves, 2019), in mind, the aim was to explore whether state change can be represented in the absence of an object. In other words, can we conceptualise a chopping event even if we do not know what is being chopped?

In Chapter 4, I explored how grammatical aspect might affect memory for events. In line with suggestions for future work by Salomon-Amend et al. (2017), I contrasted change-of-state events (e.g., to chop an onion) with events where the object remains unchanged (e.g., to type on a keyboard). The literature on aspect and event memory is sparse and only beginning to emerge. In addition, it must be noted that memory can be affected by language in two ways, which were systematically addressed in the two experiments outlined in this chapter. In Experiment 1, I studied whether language (here: a SVO sentence with grammatical aspect manipulated; "*She chopped/was chopping an onion.*") can affect event memory by guiding people's attention to specific event features prior to viewing a video clip of an event. In Experiment 2, I used the same sentence and video stimuli, but sentences were presented only after the event had been viewed.

In Chapter 5, I turned to motion events, for which aspect has previously been found to affect eye gaze to paths and destinations along trajectories. When grammatical aspect marking allows for the expression of ongoingness, speakers have been found to attend more to the trajectory itself and describe events omitting the endpoint. Speakers of non-aspectual languages, on the other end, have been found to show an ‘endpoint bias’. I extended the work by von Stutterheim et al., (2012) and Flecken et al. (2014), by testing whether previous findings generalise to everyday first-person perception of the world. Using a virtual reality (VR) setup, which allows for a dynamic and immersive 3-D setting, I tested native speakers of an aspectual language (English) and speakers of a non-aspectual language (German). Participants were walking on a treadmill across different virtual terrains encountering objects on the roadside and at the end of the road while their eye gaze was recorded. Using VR allowed me to study whether results from previous studies on linguistic experience and event cognition would replicate to situations reflecting more ecologically valid settings.

In Chapter 6, I have summarised the results from the experimental chapters. I have discussed commonalities and differences of the observed patterns of results across all chapters, and outlined future directions of research on aspect and event cognition.

6. 1. Methodological Considerations

The approaches in this thesis were chosen in consideration of existing work from two fields: event cognition theory and psycholinguistic work on language and thought. In doing so, this thesis aims to (1) add to a growing body of research that suggests that language, and specifically aspect, can affect event model construction, while (2) taking into account theoretical accounts of event cognition, and specifically IOH (Altmann & Ekves, 2019). Using a variety of manipulations and methods (e.g., looking at different types of events, using behavioural and online measures, and using verbal and nonverbal tasks) enabled me to go beyond studying whether effects of language occur, instead focusing on how and under which conditions such effects arise.

CHAPTER 2

THE STATE OF THE ONION: GRAMMATICAL ASPECT MODULATES OBJECT REPRESENTATION DURING EVENT COMPREHENSION[4]

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.

[4]This chapter was published as: Misersky, J., Slivac, K., Hagoort, P., & Flecken, M. (2021). The state of the onion: Grammatical aspect modulates object representation during event comprehension. *Cognition, 214*, 104744.

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’ 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 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 (2019) 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 tokenisation of the object and other event entities. Altmann and Ekves (2019) 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 (Radvansky & Zacks, 2014; Altmann & Ekves, 2019). The ways in which time is encoded linguistically, however, differs across languages (Klein, 2009). 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 progressive 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 defocuses 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, 2009). 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 (i.e., 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 toward 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 progressive), 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 magneto-encephalogram (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 (Zacks, 2020; Amoruso, Gelormini, Aboitiz, Alvarez González, Manes, Cardona, & Ibanez, 2013). 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 and colleagues (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, Stanfield & Yaxley, 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 (Magliano & Schleich, 2000; Becker, Ferretti, & Madden-Lombardi, 2013) 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 focusing 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., Kang et al., 2019; Hindy et al., 2012) 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, characterised 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 characterised both the objects' original (NSC) and their changed, resultant states (SC) (cf. Kang et al., 2019; Hindy et al., 2012; 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 focused 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 characterised 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

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.

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.

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).

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: $\beta = 2.176$, $t(122.710) = 20.74$, $p < .001$; SC items: $\beta = 2.673$, $t(172.820) = 37.33$, $p < .001$), 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

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.

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).

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.

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. Main Experiment

Participants

Based on previous ERP research using the sentence-picture verification task with 18-20 participants (Hirschfeld et al., 2011; 2012), 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[5]. 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.

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

---

[5]Two participants were excluded because of data recording problems, two further participants had only 70 percent 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).

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 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 leeks*" and "*was chopping the onion*"). U pictures were 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.

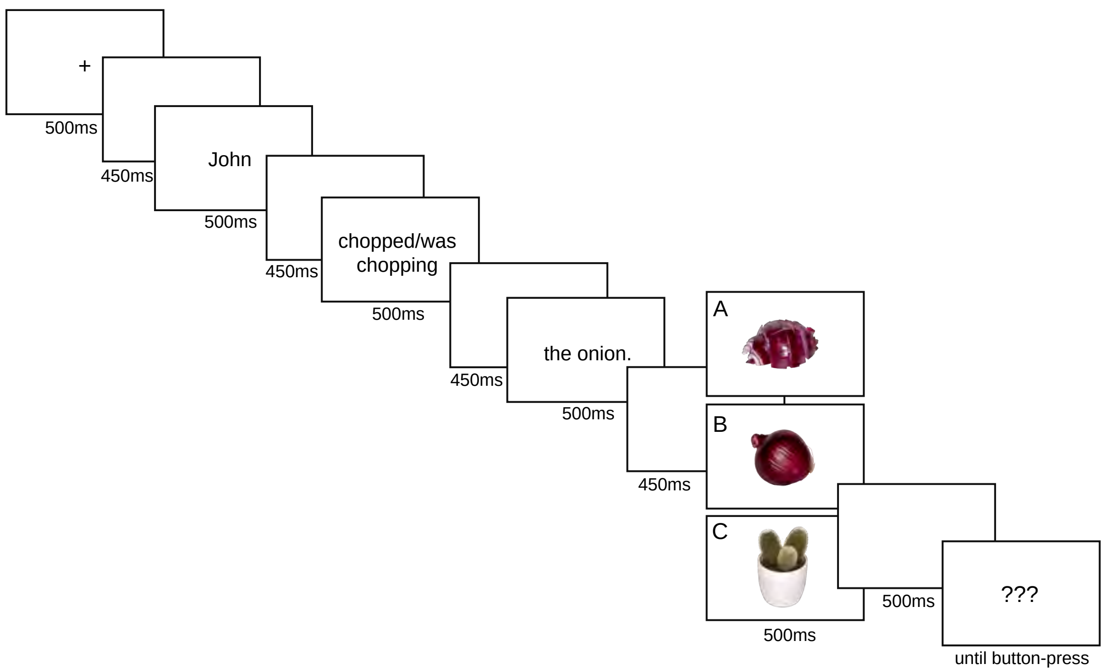


Figure 1: *Procedural EEG setup with all possible Picture manipulations; A showing the state change (SC), B showing the no state change (NSC) and C showing the unrelated (U) picture.*

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 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-inch 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.

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 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 normalised 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 than 70 percent 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.

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, Brockhoff, & Christensen, 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.

	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

Figure 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.*

Both the hypothesis-driven and the whole-head analyses followed the same steps when setting up the linear mixed effect model analysis. A 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 to 500 millisecond time-window consisting only of an intercept and a (centred and scaled) continuous predictor of the pre-stimulus baseline (GLM-based baseline correction, Alday 2019) in the fixed effects. There is no predictor 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 (centred and scaled) continuous predictors for pre-stimulus baseline and the predictor derived from the first step ("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 (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 ($\beta = 0.063$, $t(2876) = 2.931$, $p < .01$) 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 a 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 characterised by a positivity in the same time-window.

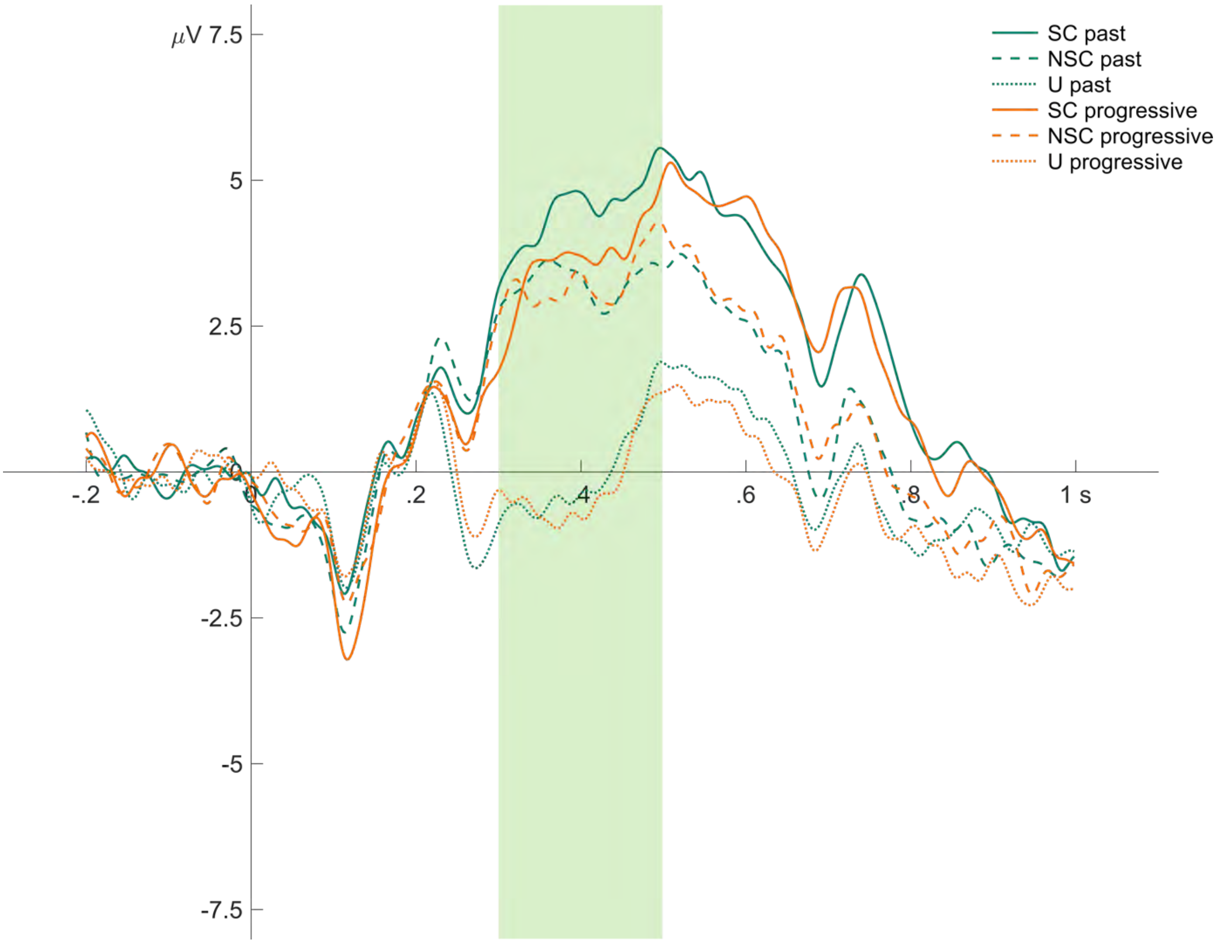


Figure 3: *Grand average waveforms collapsed across all six electrodes (Cz, CP1, CP2, P3/z/4) in the 300 to 500 milliseconds time-window (highlighted).*

Table 1. N400 ROI analysis: Output of the model on ERP amplitude in the 300 to 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]	0.031	0.0197	1.5706

***$p < 0.001$; **$p < 0.001$; *$p < 0.05$

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 to 500 milliseconds 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 positive ERP response, compared to SC pictures following progressive sentences (orange solid line).

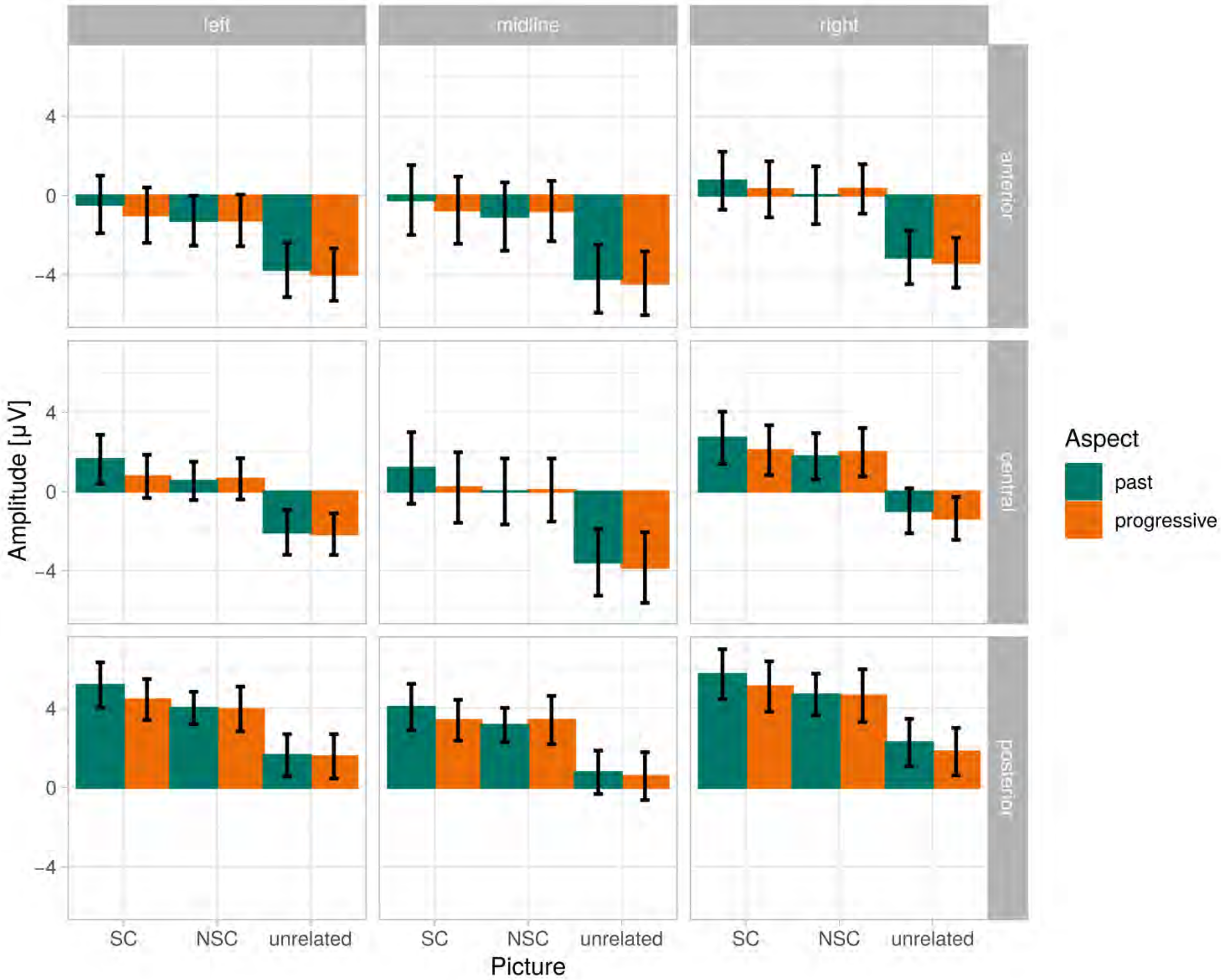


Figure 4: *Bar charts of the grand average amplitudes in the 300 to 500 milliseconds time-window for the electrodes grouped according to the topographical factors Anteriority (anterior, central, posterior) and Laterality (left, midline, right) ± 1 SEM.*

This pattern is also displayed in all panels of Fig. 4, highlighting more positive responses for SC pictures overall, and differential processing of SC pictures dependent on 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).

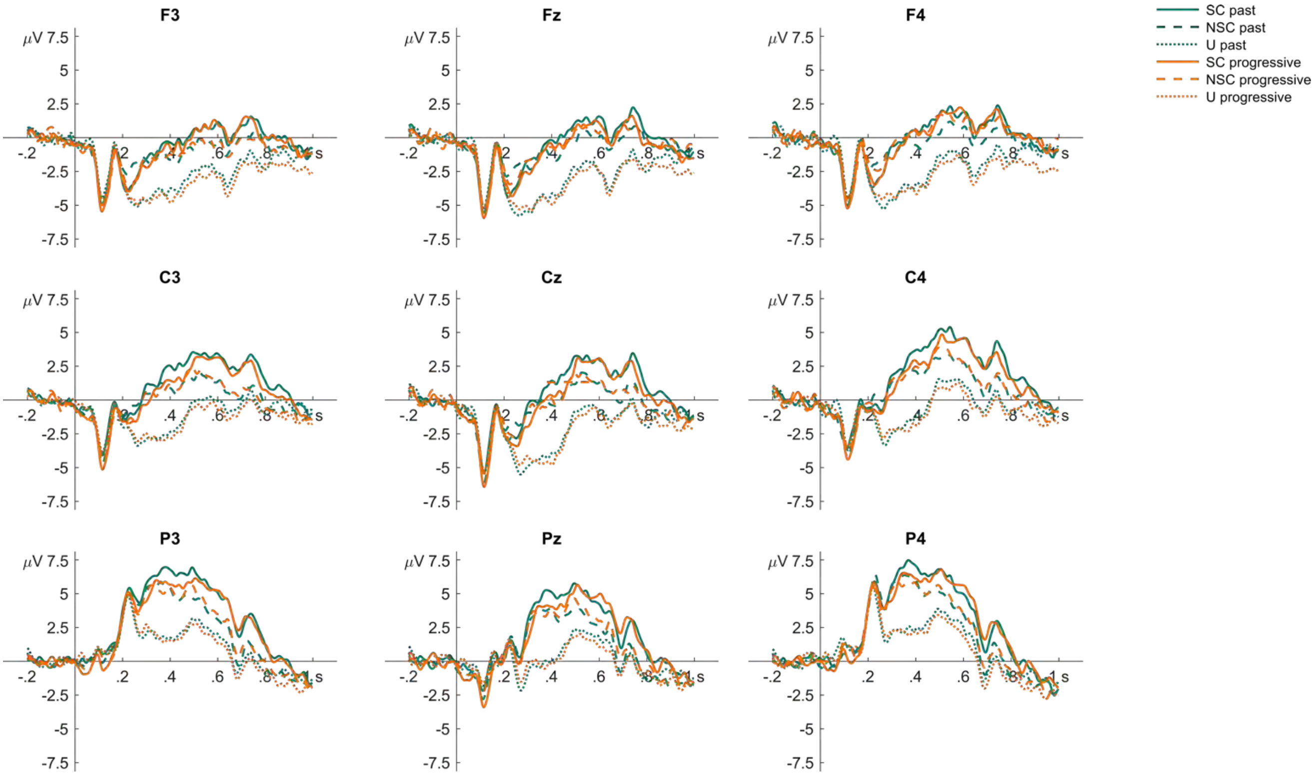


Figure 5: *Grand average waveforms at nine representative electrodes in frontal (F3/z/4), central (C3/z/4) and posterior (P3/z/4) areas.*

Table 2. Whole-head analysis: Output of the model on ERP amplitude in the 300 to 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	0.031	8.6497***
Picture [NSC]: Aspect [perfective]	-0.0094	0.0065	-1.4578
Picture [SC]: Aspect [perfective]	0.021	0.0065	3.2244**
Picture [NSC]: Anteriority [post]	-0.0049	0.0065	-0.7539
Picture [NSC]: Anteriority [ant]	0.0039	0.0065	0.601
Aspect [perfective]: Anteriority [post]	-0.0036	0.0065	0.5607
Aspect [perfective]: Anteriority [ant]	0.004	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]	0.032	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	-0.212
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]	-0.002	0.0092	-0.222
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	0.141
Picture [NSC]: Aspect [perfective]: Anteriority [post]: Laterality [right]	0.0037	0.0092	0.403
Picture [NSC]: Aspect [perfective]: Anteriority [ant]: Laterality [right]	-0.0039	0.0092	-0.429

***$p < 0.001$; **$p < 0.001$; *$p < 0.05$

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 focused 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 and colleagues' work (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, 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; Santín et al., 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 of 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 (2020), 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, 2009) 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 conceptualisations 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.

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.

5. 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 85 objects, see table below).

Past progressive	Simple past	Object	Pool of unrelated object pictures
was beating	beat	the cream. the butter.	lego tower flower
was biting	bit	into the sandwich. into the cookie.	scarf cup
was boiling	boiled	the pasta. the water.	paper plane necklace
was burning	burned	the toast. a match.	nail coca cola
was carving	carved	the pumpkin. wood.	schnitzel bubble gum
was chopping	chopped	the onion. the leek.	sentence shoes
was cracking	cracked	the chocolate bunny. the walnuts.	box poem
was crushing	crushed	the pepper. the ice cubes.	brownies bin
was cubing	cubed	the cucumber. the cheese.	lollipop package
was dicing	diced	the tomatoes. the garlic.	doll barbie
was dipping	dipped	the crackers. the bread.	email book
was dissolving	dissolved	the powder. the pill.	backpack puppet
was frying	fried	the mushrooms. the steak.	frisbee puzzle
was grating	grated	the parmesan. the carrot.	chips cushion
was grinding	ground	chillies. coffee.	socks tennis ball
was halving	halved	the avocado. the melon.	dough pencil
was juicing	juiced	grapefruits. oranges.	coins clothes
was mashing	mashed	the peas. the beans.	keys fork
was melting	melted	the ice cream. the chocolate.	shopping list soup
was mincing	minced	parsley. almonds.	pear cactus
was peeling	peeled	the banana. the mandarin.	ring magazine
was roasting	roasted	the chicken. the peppers.	card house house
was skinning	skinned	the potato. the apple.	hat pot

was slicing	sliced	the bagel. the pizza.
was squeezing	squeezed	the limes. the lemons.
was breaking	broke	the vase. the clock.
was coloring	colored	the drawing. the painting.
was destroying	destroyed	the files. the notes.
was lighting	lit	the cigarette. the candle.
was scratching	scratched	the screen. the phone.
was spilling	spilled	the paint. the whiskey.
was staining	stained	the bib. the blanket.
was tearing	tore	the photo. the shirt.
was shredding	shredded	the contract. the bank statements.

napkin
bracelet
dowl
milk
waffle
candy
essay
knot
lamp
postcard
hummus
lasagna
bucket
bonbon
plastic bag
teddy bear
balloon
text message
newspaper
handbag
dice
rock
crossword
pudding
skirt
tights
boomerang
playdough
crayon
marbles
toys
wallet
spoon
salad
smoothie
plant
juice
cherries
catalogue

CHAPTER 3

A CHOPPED ONION IS A BIT LIKE CHOPPED LEEKS: VERB SEMANTICS AND ASPECT CONSTRAIN SITUATION MODELS DURING EVENT COMPREHENSION

Abstract

In two web-based experiments we tested how verb semantics and grammatical aspect affect event conceptualisation. In Experiment 1, and building on Misersky et al. (2021, Chapter 2), a sentence-picture verification task was used. Participants read SVO sentences, where verb semantics suggested a change-of-state in an object (e.g., "*John chopped/was chopping the onion.*"). Following the sentence, participants saw a picture of the object in its original unchanged state (NSC) or in its changed state (SC) or the picture of an unrelated object (U). For each picture, they had to indicate whether the object was mentioned in the sentence. Reaction times were recorded from picture onset. No effect of Aspect was observed. A main effect of Picture revealed that SC pictures were overall responded to fastest suggesting participants simulated the event during sentence comprehension, which led to the activation of a specific object state. In Experiment 2, a similar setup was used. However, in the sentences the object noun was decisively omitted (e.g., "*John chopped/was chopping the...*") to test whether people represent state change even if the object is unknown. Participants' task was to indicate whether the picture was a good fit given the sentence. Again, SC pictures were responded to fastest, suggesting a state change can be represented in limited sentence contexts. An interaction of Picture type and Aspect showed that NSC pictures were responded to faster after perfective compared to progressive sentences.

1. Introduction

In recent years, the field of event cognition has gained more attention in cognitive science research, as it is concerned with how we conceptualise our everyday experiences in the interface with action, vision, and language (e.g., Zacks, 2020). One of the pillars of any event is the dimension of time, which together with space, frames the event and gives it structure. In addition, the event is characterised by its participants, such as the people and objects in it, and their relationship to one another. In this project, we build on previous work, where we showed that the expression of time in language modulates our understanding of events (Misersky, Slivac, Hagoort, & Flecken, 2021).

Generally speaking, our understanding of an event is guided by prior experiences of similar nature. These prior experiences are captured in memory representations of events, called event models. Event models (akin to situation models; van Dijk & Kintsch, 1983) guide our online processing of events as they unfold by informing a working model of a current experience. These working models take into account knowledge of the spatio-temporal framework of an event, the event participants, and so forth, in order to segment the ongoing activity into discrete events with beginning and end (so-called event boundaries). While Event Segmentation Theory (EST; Zacks & Swallow, 2007; Zacks & Tversky, 2001) describes this 'chunking' mechanism, more recently, theoretical contributions have focused also on how we process the inner structure of events. Specifically, the Intersecting Object Histories account (IOH; Altmann & Ekves, 2019) relies on the same predictive mechanisms as outlined in EST, yet is concerned with what happens within the boundaries of an event. According to Altmann and Ekves (2019), objects are at the core of event understanding. We track objects and how they change states as the event unfolds. The relation of the object to other event participants over the time course of the event helps us conceptualise the event as a whole. Importantly, both EST and IOH acknowledge that it is not the direct (perceptual) experience alone which informs our event models. Additionally, we also experience events indirectly through language, for example, when reading the news or talking about something we did the other day.

Previous research has shown that we indeed construct event models based on linguistic input, and that we do so in an anticipatory fashion. Employing the visual world paradigm (VWP), Altmann and Kamide (1999) presented participants with visual displays whilst playing auditory sentence stimuli and tracking participants' eye gaze. One of their stimuli showed a scene with a boy, a cake, and some other objects. Upon viewing the scene, participants would either hear "*The boy will <u>move</u> the cake.*" or "*The boy will <u>eat</u> the cake.*" Since the cake was the only edible object in the scene, participants directed their gaze to the cake much earlier in the "*eat*"-condition compared to the "*move*"-condition. This example shows that verb semantics rapidly constrain participants' working models of an event, in this case, by selecting the most likely upcoming referent.

Not only the most suitable referent object in an event is constrained by a verb; verbs have also been shown to convey what an object might look like. In a recent study, also using the VWP, Kang, Joergensen and Altmann (2020), have shown that sentences where the verb suggested state change in an object (e.g., "*The chef will chop the onion.*") guided participants eye movements such that their gaze was directed at the picture showing the corresponding state (a chopped onion) even when one of the competitor objects was the same object in the original, unchanged state (an unchopped, round onion). In a previous study, Kang, Eerland, Joergensen, Zwaan and Altmann (2019) studied the role of verbs in conceptualising object states using a sentence-picture verification task. In Experiment 2, they presented participants with sentences in which the state of an object (e.g., an ice cream) had either undergone substantial change ("*She <u>dropped</u> the ice cream.*", suggesting the ice cream was no longer intact) or had undergone minimal change ("*She held the ice cream.*", suggesting the ice cream was in its original state). Thereafter participants saw a picture of an ice cream in the cone (original state) or on the floor (changed state) or a filler object, and had to indicate whether the object in the picture was mentioned in the sentence. Participants were faster to give a correct response when there was a match of the state of the ice cream suggested in the sentence and the state shown in the picture. Together, these results indicate that participants used the information in the verb in combination with the object noun to create a working model, which included information on the state of the affected object, which in turn informed their visual attention and response speed.

Akin to the results by Kang et al. (2019), there have been studies using the sentence-picture verification task to study whether we simulate specific features of objects, such as their orientation, shape, and colour. For example, Zwaan and Pecher (2012) aimed to replicate six behavioural studies using an online version of the sentence-picture verification task, in which 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 sentences guided the participants' responses to the objects in the pictures. Zwaan and Pecher (2012) consider the match advantage an automatic and effortless process, in part afforded by relevance of orientation and shape to the interpretation of the event as described in the sentence. Accordingly, people seem to infer the trajectory of the action relevant to affording the shape or orientation of an object.

Taken together, the behavioural research on language and event model construction, focusing on conceptualisation of affected objects, their shape and state change, suggests that language plays an important role in shaping and guiding our conceptualisation of events. Importantly, the results obtained by Kang et al. (2019) with sentences in the past tense (Experiment 2) did not hold when participants read the same sentences in the future tense ("*She will drop/hold the ice cream.*", Experiment 3). For past tense sentences, the original state pictures were verified faster after sentences suggesting minimal change, whereas changed-object pictures were verified faster after sentences suggesting substantial change. For future sentences, however, only the changed state picture was verified more quickly after sentences suggesting substantial change but there was no match advantage for original state pictures. These results highlight that the semantics of verbs convey state change and that comprehenders use this information to form visual representations of the objects in events. Further, they suggest that time in language may play a particularly important role in event understanding. Specifically, verb tense guides comprehenders' inferences about what objects in events look like, at different phases of the overall time span of the event: Events described in the past allow for constructing a holistic event model, including the changed state of the object (Altmann & Ekves, 2019).

Recall that time is a core feature in event cognition, which provides an event with structure. However, the ways in which time is encoded linguistically differs across languages (Klein, 2004). English, the language this study focuses on, expresses time in part through tense and aspect, with verbs generally describing a situation as relative to time (Klein, 1994). The expression of time at the verb level can occur through grammatical features. Simplified it can be said that verb tense acts as a reference to the time of utterance; it links the time of utterance to what is called the Topic Time, which relates to 'the time for which the particular utterance makes an assertion' (Klein, 1994, p. 37). Grammatical aspect marks different viewpoints on the internal temporal structure of the same event, such as completed or ongoing. As a result, aspect provides information as to what time interval or phase of the entire situation is under discussion (Klein, 1994); in other words, it specifies the Topic Time. Tense and aspect function independently 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 example, the simple past tense, signalling the perfective aspect in English, marks the situation as completed, i.e., the Topic Time covers the entire time-window associated with the situation 'to chop an onion'. The second sentence is marked for ongoingness (progressive aspect), such that the Topic Time is embedded within the overall time span of the situation 'to chop an onion'. Definitions of tense, aspect, and combinations of the two pose a matter of discussion, for example, when verbs describe states (Klein, 2004). However, verbs describing irreversible action (*accomplishments*, Vendler, 1967; e.g., to chop) in the past tense, can distinctly mark ongoing and complete events in English for the most part. Aspect may thus allow us to interpret an event from distinct viewpoints.

Radvansky and Zacks (2014) acknowledge the potential relevance of grammatical aspect to the processes important to event comprehension. Aspect marking may be particularly relevant to events during which an object undergoes change, as highlighted by Altmann and Evkes (2019). Here, grammatical aspect could be considered to cue for a specific object state, reflecting a given phase of an event. This way, aspect can give us detailed information about the event content, such as what objects look like in different phases of the event, by honing in on the event's inner temporal structure. Time is central to event cognition theory, and linguistic features such as aspect may affect our understanding of events by cueing for a specific event phase, marked by specific object states.

The existing literature on grammatical aspect and event cognition has for the most part used behavioural measures. Generally speaking, the experimental manipulation of verb aspect has shown that the (English) progressive, focusing on an event's ongoingness and defocusing event boundaries, affects how we understand events. In their Experiment 2, Madden and Zwaan (2003) 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 completed event pictures compared to ongoing event pictures 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 (in this case, ongoing event pictures) and mismatching (completed event) pictures. The perfective constrained the event model such that the focus was on the resultant state, leaving only the completed picture considered as a match. The progressive, however, foregrounded a variety of features relevant to the various stages of the event, leading to inconsistencies between participants regarding their ongoing event representations. So, while perfective trials resulted in a match benefit—i.e., a strong preference for the completion state—, this was not the case for progressive trials.

More recently, the effects of grammatical aspect on event cognition have also been studied using event-related potentials (ERPs), to examine how language affects the processing of change-of-state events prior to decision-making. In a sentence-picture verification study, adapted for electroencephalogram (EEG), Misersky et al. (2021), presented participants with English sentences in which all verbs suggested irreversible change-of-state in an object (e.g., to chop an onion). The sentence was presented 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), no state change (NSC; an onion in its original, unchopped state), or an unrelated object (U; a cactus, control condition). ERPs were recorded from picture onset, with a focus on ERPs in the 300 to 500 milliseconds time-window. Given the semantics of the verbs, a general cognitive bias towards action goals and state change (e.g., Lakusta & Carey, 2015), as well as all described events taking place in the past, a main effect of Picture was expected. Specifically, SC pictures were expected to be more compatible with the event model compared to NSC pictures, which was indeed observed in the ERP data. In line with the main research question, the authors also expected an interaction of Picture and Aspect. The perfective was expected to cue for an object in one specific (resultant) state (i.e., the SC pictures), the progressive was expected to defocus the completion state of the object. The ERPs indeed revealed an interaction: For sentences in the perfective ("*chopped*"), participants showed a larger P300 response for SC pictures compared to NSC pictures. The perfective sentences cued for a specific object state representation, namely, that of an object in its changed resultant state. In line with functional interpretations of the P300 (Kok, 2001), the higher amplitude for SC pictures compared to NSC pictures was taken to reflect a better perceived match between sentence and picture.

Taken together, the existing research can be summed up as follows: First, and broadly speaking, language can constrain event models with regards to informing predictions about likely object referents ("*The boy will <u>eat/move</u> the cake.*", Altmann & Kamide, 1999). Second, the semantics of sentences can cue for a specific visual representation of an object (e.g., the shape effect; "*John put the pencil <u>in the drawer/in the cup</u>.*", Zwaan & Pecher, 2012), which may also encompass the inference of a specific action trajectory (e.g., how the pencil is held when put in the cup vs the drawer). Further, the combination of (change-of-state) verb and direct object noun can specifically cue for object state ("*The chef weighed/chopped the onion.*", Kang et al., 2019). Third, grammatical aspect marking—at the verb and in combination with an object noun—, plays a role in which object state is represented ("*The chef chopped/was chopping the onion.*", Misersky et al., 2021). Specifically, the perfective, particularly in the past, cues for the changed state of an object as shown by online (Misersky et al., 2021) as well as offline measures (e.g., Madden & Zwaan, 2003; Madden-Lombardi, Dominey, & Ventre-Dominey, 2017).

1. 1. The Present Study

In this study, we aim to test the effect of grammatical aspect on event conceptualisation in two experiments. In both experiments, we opted to exclusively focus on change-of-state-events described in the past and use pictures of objects in original (NSC) and changed (SC) state as well as unrelated (U) pictures. Experiment 1 adapted the design used by Misersky et al. (2021) for a web-based sentence-picture verification task measuring reaction times, where participants were presented with a sentence in the past (SVO, e.g., "*John chopped/was chopping the onion.*"). Experiment 1 thus served as an extension of Misersky et al. (2021) to gain a clearer insight into whether the effect of aspect marking on object representation can be captured by reaction times (RTs) as a measure of conscious decision-making. In line with the ERP results, we expected a main effect of Picture, such that SC pictures would be responded to faster compared to NSC pictures. In addition, we also expected an interaction of Picture and Aspect, where SC pictures would be favoured after sentences in the perfective compared to the progressive. Experiment 2 used a comparable web-based setup and the same stimuli as Experiment 1, yet with the object in the sentence presented solely by a picture instead of a noun (cf. Wicha, Moreno, & Kutas, 2003). Object characteristics, including their states, can be inferred through the semantics of the entire sentence (e.g., Zwaan & Pecher, 2012), yet recent work has shown that the verb may be enough to successfully build a model of a change-of-state event (e.g., Kang et al., 2019, 2020; Misersky et al., 2021). In addition, Wicha et al. (2003)

show that comprehenders can predict an upcoming direct object referent in sentences based on preceding sentence material (including the verb plus the gender-marked article of the noun). Experiment 2 was set up to study if we conceptualise object states in the absence of a full event context and, if so, whether this is modulated by aspect marking. To this end, the linguistic input describing the event in Experiment 2 was reduced; the event model to be constructed by a comprehender was thus mainly driven by the inferred trajectory of the action carried out by an agent, and associated object affordances, encoded in the verb and the aspect marking on the verb. The aim was to study whether we can represent state change events, including the visual representation of the state changed object, even if the specific object undergoing the change is unknown. Note that for Experiment 2 we were unable to run a sentence-picture verification task, since the picture was part of the sentence. Instead, the task was adapted to ask for goodness-of-fit between sentence and picture (see below). If the verb is enough of a 'driving force' to activate object states during event comprehension, we should expect differential responses to pictures showing the object in its original state (NSC) compared to its changed state (SC). Specifically, responses should be faster for SC pictures compared to NSC pictures if state change is activated by the verb semantics. The role of aspect in a limited sentence context remains to be studied. However, if aspect has an effect of event conceptualisation in a limited context, SC pictures are likely to receive faster responses after sentences in the perfective compared to sentences in the progressive.

2. Experiment 1: Can Grammatical Aspect Modulate Event Conceptualisation in a Full Sentence Context?

2. 1. Methods

For the pre-test of the stimuli items, please refer to Chapter 2 (Misersky et al., 2021).

Participants

One-hundred-and-fifty native speakers of English (102 female, 48 male, 18-42 years of age, *M* = 23.59, *SD* = 5.99), participated in this study[6]. The study 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 monetary compensation for their time (6 GBP per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, The Netherlands.

Materials and Design

For the materials, we used the same sentences and pictures as in Misersky et al. (2021). All sentences were in the format SVO (e.g., "*John chopped/was chopping the onion.*"), with all verbs suggesting a change-of-state (e.g., to chop). Within each sentence, we manipulated the aspect of the verb as perfective (e.g., "*chopped*") and progressive (e.g., "*was chopping*"). All sentences were followed by either a picture of the object having undergone a state change (SC; a chopped onion), no state change (NSC; an onion in its original, unchopped state), or an unrelated object (U; a cactus, control condition). We opted for ten unique lists from Misersky et

[6]Forty-one participants were excluded from the final analysis: Two due to being non-native speakers of English, three due to incomplete data sets, five due to technical issues, another two due to a high number of missed (timed-out) trials, and 29 due to their low accuracy (< 75% correct) in any of the conditions. The analysis of the data is hence based on 109 data sets.

al. (2021), which were distributed to batches of 15 participants. Item presentation for each list was randomised per participant.

Procedure

The study employed a sentence-picture verification task: First, participants read a sentence presented as a whole, and were instructed to press the spacebar when they had finished reading to reveal the picture. Like in Misersky et al. (2021), the picture was flashed for 500 milliseconds. The task was to indicate via Yes/No button press, whether the object in the picture was mentioned in the previous sentence. Participants were told to do so as soon as possible from picture onset. Participants received four practice trials to familiarise themselves with the task. Reaction times were recorded upon picture onset. Trials timed out after 2000 milliseconds.

Data Analysis

For the analysis, we implemented a low cut-off for the reaction times (RTs) and thus trimmed the RT data at 300 milliseconds. However, no data needed removing. Due to the built-in timeout of the experiment, we did not use an additional high cut-off. The RT data were log-transformed, and analysis was carried out only on correct responses. We opted to analyse the data using linear mixed effect models in R (Version 3.6.3) with lme4 (Version 1.1-26, Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (Version 3.1-3, Kusnetzova, Brockhoff, & Christensen, 2017) packages. The fixed effects factor Picture was sum-coded (SC = 1, NSC = -1, U = 0, and SC = 1, NSC = 0, U = -1). For the random effects (RE) structure, we started out with intercept-only effects for Participant. Goodness-of-fit of the resulting models was assessed using the rePCA function (lme4, Bates et al., 2015), and the RE structure was adjusted and re-assessed according to this same method until model fit was optimised. T-tests used the Satterthwaite's method.

2. 2. Results

Table 1. RT analysis: Output of the model on log RTs of the correct responses to the pictures (lmer (logRT ~ 1 + Picture * Aspect - Aspect + (Picture | Participant))

Fixed Effects	β Estimate	SE	*t*-Value
Intercept	6.503	0.016	415.127***
Picture SC	0.029	0.005	5.820***
Picture NSC	-0.027	0.007	-3.868***
Picture NSC : Aspect prog	0.004	0.007	0.624
Picture SC : Aspect prog	0.001	0.007	0.158
Picture U : Aspect prog	0.009	0.007	1.391

****p* < 0.001; ***p* < 0.001; **p* < 0.05

The results indicate a main effect of Picture: As expected, SC pictures were responded to fastest (*M* = 690.601, *SD* = 133.824). U pictures received overall slowest responses (*M* = 724.948, *SD* = 127.049), placing responses to NSC pictures in between (*M* = 720.364, *SD* = 142.782). Unlike hypothesised, no interaction effects with Aspect were observed.

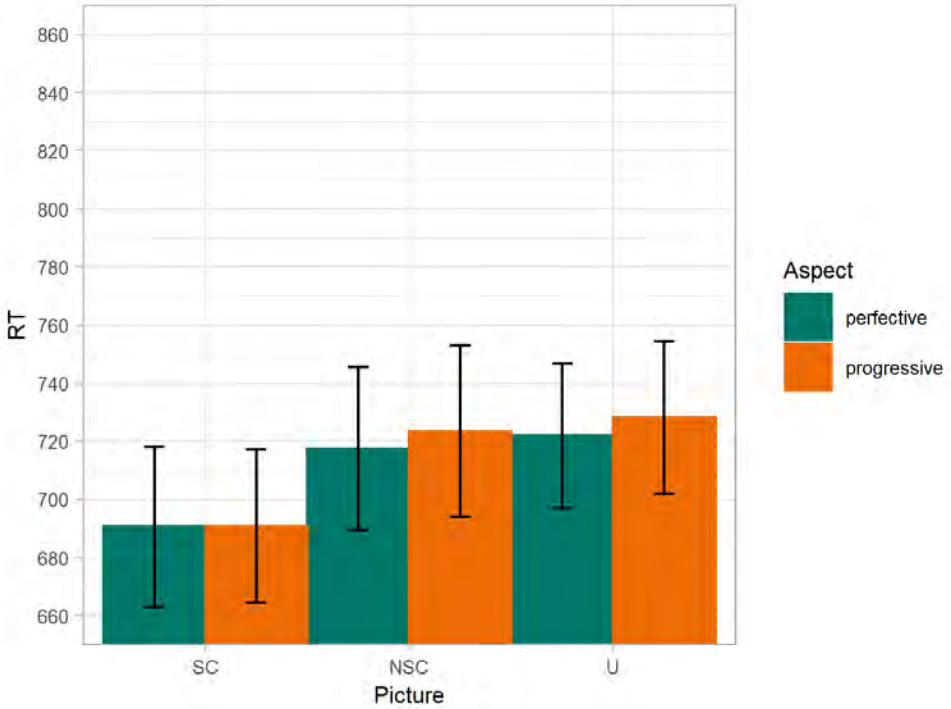


Figure 1: *Bar charts of the average reaction times (RTs, in milliseconds) for each Picture type (SC, NSC, and U) and Aspect (perfective and progressive) ± 1 SEM.*

The results point towards the verb semantics having activated a particular visual representation in line with the state change of the object as shown in the SC pictures. Building on this, in Experiment 2, we explored if we conceptualise state change in the absence of a full event context and, if so, whether this is modulated by aspect marking.

3. Experiment 2: Can Grammatical Aspect Modulate Event Conceptualisation in a Limited Sentence Context?

3. 1. Methods

3. 1. 1. Pre-Test of Stimuli Items

For the main experiment we used sentences with state change verbs, describing an action which would irreversibly change the physical appearance of an object. Sentences all followed SVO sentence order, yet the object noun itself was replaced by a picture of the object, either in its changed state (SC), in its original state (NSC) or an object semantically unrelated to the verb (U). In two pre-tests we assessed recognisability of the objects, as well as acceptability of verb-object pairs.

3. 1. 1. 1. Pre-Test 1: Fill-in-the-Blanks Task

Participants

Forty native speakers of English (24 female, 14 male, 2 non-binary, 18-39 years of age, *M* = 22.12, *SD* = 5.0), participated in this pre-test. The pre-test 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 monetary compensation for their time (6 GBP per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, The Netherlands.

Materials and Design

Since the object picture in the main experiment would appear for a brief amount of time only (500 milliseconds), we used this pre-test to test whether participants were able to recognise and name the objects. Participants rated 64 NSC pictures as well as the corresponding 64 SC pictures, and 64 U pictures, making a total of 192 items.

Procedure

Participants were instructed to name pictures of objects in a fill-in-the-blanks task. On each trial, participants read "*This is...*" (for NSC and U pictures, *N* = 128) or "*This is a* [adverb, e.g., *chopped*]*…*" (for SC pictures, *N* = 64). After reading the description fully, they had to click to reveal the image of the object. The image appeared for 500 milliseconds and after disappearing, a blank box appeared in which the participants could type in the name of the object. Clicking Enter started the next trial. If no answer was provided, the next trial started automatically after 90 seconds. The pre-test started with five practice trials. All trials were randomised for each participant.

Data Analysis

First, we checked the list of individual responses, correcting and standardising spelling. In cases where participants provided adjectives or other descriptors (e.g., "*a piece of bread*", "*red apple*"), we reduced their responses to only the head nouns. Further, we opted to standardise items such that only the singular or plural were used consistently for the same item. Then, we split by item to get the unique responses for each picture to calculate Shannon's entropy to evaluate the name agreement for the items. Based on the distribution, we opted for a cut-off of 3. This affected twelve of the SC and NSC pictures, and three of the U pictures. Since in the main experiment, each verb was supposed to occur with two pictures of each condition (for SC, NSC, and U conditions, respectively), high entropy (i.e., low name agreement) for any picture could result in a substantial loss of items. For the SC and NSC pictures, we then checked whether the labels participants provided semantically suited the verb we planned to pair it with. We then calculated the percentage of Yes (i.e., picture is acceptable with regards to verb) responses and were able to keep six of the twelve SC/NSC pictures due to a score higher than 75% Yes (is acceptable). For the three U pictures, we opted to replace them with pictures with higher agreement. To decide which new verb and object combinations were most acceptable, and in line with the Acceptability judgment pre-test (see below), we additionally ran a small-scale pre-test asking participants (*N* = 22) to rank six objects as worst and least bad fits with the suggested verbs, then opted for the verb and object combinations that got the most worst-fit responses (e.g., "*to squeeze a cactus*").

3. 1. 1. 2. Pre-Test 2: Acceptability Judgment Task

Participants

Forty native speakers of English (29 female, 11 male, 18-42 years of age, *M* = 22.5, *SD* = 6.32), participated in this pre-test. The pre-test 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 monetary compensation for their time (6 GBP per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, The Netherlands.

Materials and Design
In this pre-test we tested verb-object compounds in an acceptability judgment task to ensure that our choice of SC/NSC and U items was justified, and indeed reflected (un)familiar events. Participants read each verb (*N* = 32) four times, each time paired with a different object. Two of the objects were semantically related to the action verb, whereas two were not. In total, this resulted in 128 ratings per participant.

Procedure
In each trial, participants read a phrase (e.g., "*to chop the onion*"; "*to fry the frisbee*") and were asked to indicate how well the verb and object go together, which they had to rate on a Likert scale ranging from 1 (= do not go together at all) to 7 (= go together very well). The pre-test started with five practice trials. Trials were randomised for each participant.

Data Analysis
Objects and verbs thought to be going together well indeed received high ratings (*M* = 6.15, *SD* = 1.43). Items with a mean below 4.74 (i.e., 1 SD above/below mean) were removed, which affected two items. Objects and verbs thought not to be going together well received low ratings (*M* = 1.95, *SD* = 1.38). Items with a mean above 3.35 (i.e., 1 SD above/below mean) were removed, which affected four items. In total, six verbs needed removing after this pre-test.

3. 2. Main Experiment

3. 2. 1. Methods

Participants
One-hundred-and-fifty native speakers of English (100 female, 50 male, 18-42 years of age, *M* = 24.79, *SD* = 6.2), participated in this study[7]. The experiment 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 monetary compensation for their time (6 GBP per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, The Netherlands.

Materials and Design
Based on the two pre-tests we used 24 unique verbs, across six conditions resulting in 144 items. For the materials, we used the same sentences and pictures as in Misersky et al. (2021). All verbs suggested an irreversible change-of-state (e.g., "*chop*"), and we again manipulated the aspect of the verb as perfective (e.g., "*chopped*") and progressive (e.g., "*was chopping*"). Note that in this experiment, participants were presented with only a limited sentence context. Thus, they did not receive the full sentence, but only the subject and verb followed by a definitive article (e.g., "*John chopped/was chopping the…*"). All phrases were followed by either a picture of an object having undergone a state change (SC; a chopped onion), no state change (NSC; an onion in its original, unchopped state), or an unrelated object (U; a cactus, control condition).

[7]Eighteen participants were excluded from the final analysis: One due to being a non-native speaker of English, five due to technical issues, another one due to a high number of missed (timed-out) trials, and eleven due to their low accuracy (< 75% correct) in any of the conditions. The analysis of the data is hence based on 132 data sets.

We created two lists, which were distributed to two batches of 75 participants. Item presentation for each list was randomised per participant.

Procedure
Since participants did not see a full sentence in this experiment, we were unable to run a classic sentence-picture verification task, yet the adapted task also required a matching of phrase and picture. First, participants read the phrase presented as a whole, and were instructed to press the spacebar when they had finished reading to reveal the picture. Like in Experiment 1, the picture was flashed for 500 milliseconds. In line with the pre-tests, the task was to indicate via Yes/No button press, whether the picture was a good fit given the previous phrase as soon as possible from picture onset. Participants received four practice trials to familiarise themselves with the task. Reaction times were recorded upon picture onset. Trials timed out after 2000 milliseconds (screen message "*Too slow!*") if no response was made.

Data Analysis
Like in Experiment 1, we implemented a low cut-off for the reaction times (RTs) and thus trimmed the RT data at 300 milliseconds. No data needed removing. Due to the built-in timeout of the experimental trials, we did not use an additional high cut-off. The RT data were log-transformed, and analysis was carried out only on correct responses. Similar to Experiment 1, we analysed the single-trial data using linear mixed effect models in in R (Version 3.6.3) with lme4 (Version 1.1-26, Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (Version 3.1-3, Kusnetzova, Brockhoff, & Christensen, 2017) packages. The fixed effects factor Picture was sum-coded (SC = 1, NSC = -1, U = 0, and SC = 1, NSC = 0, U = -1). Goodness-of-fit of the resulting models was then assessed using the rePCA function (lme4, Bates et al., 2015), and the RE structure was adjusted and re-assessed according to this same method until model fit was optimised. T-tests used the Satterthwaite's method.

3. 2. 2. Results

Table 2. RT analysis: Output of the model on log RTs of the correct responses to the pictures (lmer (logRT ~ 1 + Picture * Aspect - Aspect + (Picture | Participant))

Fixed Effects	β Estimate	SE	*t*-Value
Intercept	6.655	0.012	541.590***
Picture SC	0.008	0.004	1.826
Picture NSC	-0.024	0.006	-4.193***
Picture NSC : Aspect prog	0.017	0.007	2.404*
Picture SC : Aspect prog	-0.003	0.007	-0.437
Picture U : Aspect prog	-0.0004	0.007	-0.065

***$p < 0.001$; **$p < 0.001$; *$p < 0.05$

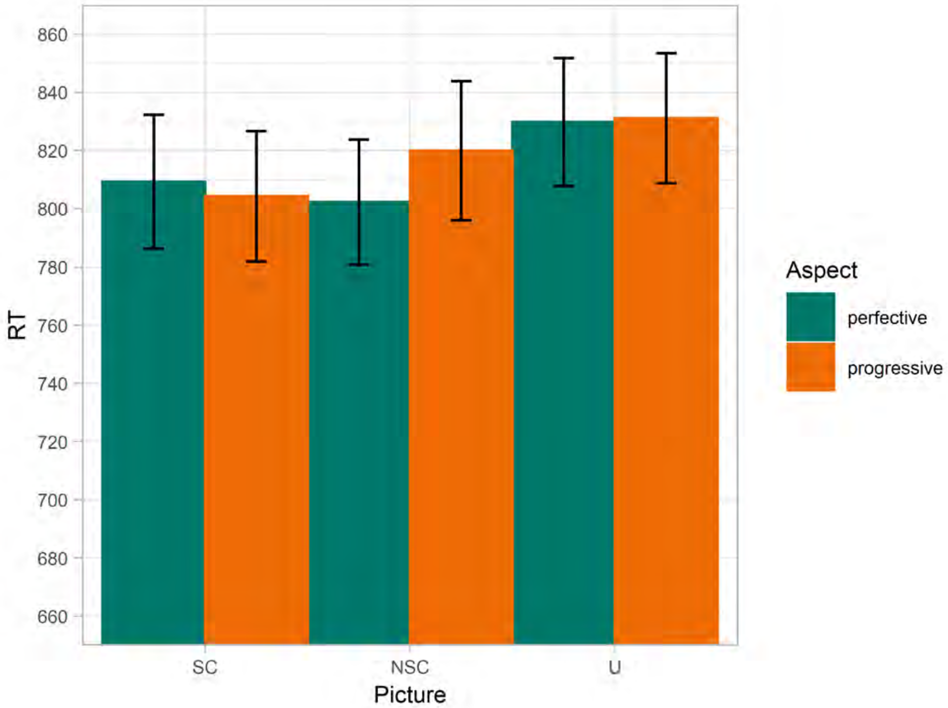


Figure 2: *Bar charts of the average reaction times (RTs, in milliseconds) for each Picture type (SC, NSC, and U) and Aspect (perfective and progressive) ± 1 SEM.*

Similar to Experiment 1, the results indicate a main effect for Picture: Overall, SC pictures were responded to fastest (*M* = 806.781, *SD* = 129.848), followed by responses to NSC pictures (*M* = 811.110, *SD* = 130.252). U pictures received overall slowest responses (*M* = 830.465, *SD* = 126.565). There was also an interaction effect; NSC pictures were responded to faster after a sentence in the perfective compared to a sentence in the progressive.

4. Discussion

The present study assessed whether grammatical aspect modulates event comprehension when reading about events. We focused on (visual) object representation after full sentences (Experiment 1) and incomplete sentences (Experiment 2). Building on previous work (Misersky et al., 2021), we used exclusively change-of-state events and presented people with objects in their original, unchanged state (NSC) or their changed, resultant state (SC). Specifically, the goal was to shed light onto the role of grammar—here: the expression of event time via aspect marking—as providing important cues to event comprehension. Previous behavioural work has shown that language can cue for specific visual representations of objects (e.g., Zwaan & Pecher, 2012). Recently, work by Kang et al. (2019) has shown that verb semantics suggesting an object state change (e.g., to drop) can affect how the object (e.g., an ice cream) is represented. In a recent EEG study, Misersky et al. (2021) have shown that after reading sentences suggesting a state change (e.g., "*John chopped/was chopping an onion.*"), pictures showing state change (SC) were interpreted as a better match than pictures showing the object in its original, unchanged state (NSC). In addition, an interaction was observed such that Aspect also constrained object representation, meaning event time had an effect on how the object in the event was conceptualised. SC pictures were considered a better match after sentences in the perfective (describing a completed event) compared to the progressive.

In the present study, and similar to Misersky et al. (2021), we also see a main effect of Picture in Experiment 1: SC pictures were responded to faster compared to NSC pictures, suggesting the verb semantics activated object properties, which is also in line with Kang et al. (2019). Unexpectedly, the interaction effect with Aspect previously observed in Misersky et al. (2021) was not present in Experiment 1. The absence of an effect of Aspect is interesting given previous findings with the same stimuli and task. Given that the specific manipulation is highly similar, it may thus be that the Aspect interaction shown in the ERP study of Misersky et al. (2021) characterises an early cognitive process, which is not captured by RT measures. Note that our picture stimuli represent the event boundaries (i.e., beginning and end states of the object undergoing change, NSC and SC respectively) rather than the intermediate and end states for which RT effects of aspect have been observed (cf., Madden & Zwaan, 2003). Though the main aim of Experiment 1 was to extend and replicate Misersky et al. (2021), follow-up work using pictures of intermediate states of object change may be a way to get additional insights into whether aspect may affect the processing of change-of-state events and object representation. To sum up, with respect to conscious decision-making as measured by RTs, in a full sentence context (verb and noun, Experiment 1), there is an advantage for SC processing, but no influence of aspect on object representation.

In one of their experiments, Kang et al. (2019) showed that when only presented with an object noun without context, participants respond fastest to pictures showing the original state of an object, comparable to our NSC pictures. This is similar to a prototype effect, where a (linguistic) label allows for the recognition of an exemplar percept (Lupyan, 2012; Attneave, 1957), suggesting state change is not represented when only an object (noun) is provided. In Experiment 2, we limited the linguistic input by doing the reverse: Instead of giving an object noun, we omitted it, and provided only a subject and verb (e.g., "*John chopped/was chopping the...*"), with the object being shown as a picture. This was to test whether an event model conceptualising a state change event can be driven by the verb (plus grammatical aspect marking on the verb) alone, even when an object is not explicitly provided. In other words, can state change be represented in the absence of an object? We again observed that SC pictures were overall responded to fastest, suggesting that even though the linguistic input was limited, the verb semantics still affected people's expectations for the specific (visual) representation of the upcoming referent, in particular, the state of an object in a change-of-state event (Kang et al., 2020). Specifically, participants seemed to have expected a changed object fitting with the semantics of the verb. We also observed an interaction effect: NSC pictures, showing the object unchanged, were responded to faster after sentences in the perfective compared to the progressive. At first sight, this effect seemed unexpected given that the perfective cues for the resultant object state (SC). Recall, though, that the perfective cues for events holistically, and in theory both beginning and end states of the event may be available, while the internal event structure may be difficult to access. By contrast, the progressive tends to defocus event boundaries by foregrounding event features associated with its internal structure (Madden & Zwaan, 2003). As a result, the beginning and end of the event (as exemplified by the NSC and SC pictures, respectively), may generally be less available and therefore responded to slower after sentences in the progressive. Further, in the case of events describing accomplishments, the progressive may lead to expectations specific to the event outcome (Baggio, van Lambalgen & Hagoort, 2008). Thus, the beginning state (NSC) in Experiment 2 may have been even less accessible after sentences in the progressive due to the verb semantics signalling a change-of-state event, and a general cognitive bias for goal states (Lakusta & Carey, 2015) further adding to the observed pattern of results.

In sum, while an effect of Aspect was observed in Misersky et al. (2021) using the same stimuli and setup as Experiment 1, the effect was absent in the latter, perhaps due to the difference in measures analysed. However, an effect of Aspect was observed in Experiment 2, which did not resemble the effect previously shown by Misersky et al. (2021). In contrast to the progressive ("*was chopping*"), the perfective ("*chopped*") leads to a holistic event representation including the beginning and end state of the object. In a full sentence context and when analysing online cognitive processing measures (Misersky et al., 2021), the effect of Aspect seems to show on the resultant state (SC), whereas in a limited context using behavioural measures this seems to show on the initial state of the object (NSC). While our results indicate that aspect modulates situation model construction, the observed pattern of results requires further study. We suggest using a comparable setup with behavioural measures and ERPs in a lab-based study to gain additional insights.

The question remains as to why the observed pattern of effects, specifically regarding the effect of Aspect, differ across the two experiments. A possible explanation may be that the tasks in the two experiments were similar, yet not identical. In Experiment 1, participants had to indicate whether the object was mentioned in the sentence. To this end, they had to explicitly match the provided object label from the sentence to the picture, similarly to other studies looking at aspect marking and event processing (e.g., Madden & Zwaan, 2003). In Experiment 2, however, using a sentence-picture verification task was not an option since we limited the sentential context to omit the object noun. Instead, we asked about the goodness-of-fit of the picture with the previous (limited) sentence. Despite having the sentence-picture pairings pre-tested, the task is by nature more prone to subjective labelling and interpretation of goodness-of-fit. Additionally, in Experiment 2, the object trajectory (cf. Altmann & Ekves, 2019) may have been generally harder to grasp, because participants were not being presented with the label for the object and were thus unable to track the object trajectory very precisely since the object was unknown. This may not have affected accuracy in the responses but RTs due to potential hesitations in answering the question in Experiment 2 compared to Experiment 1.

In the present study, we have shown that a full sentence context allows for the representation of a change-of-state event, including the state change of the object (Experiment 1). This is also true for limited sentence contexts, suggesting that a verb signalling a change in state may be enough to activate a state-specific (visual) representation of an object (Experiment 2). In line with previous studies (Kang et al., 2019, 2020; Misersky et al., 2021), verb semantics appear to be the 'driving force' behind building a model representation of a change-of-state event. In particular, the availability of the visual representation of a changed object in both full and limited sentence contexts is in line with IOH, suggesting that object states are tracked during event understanding (Altmann & Ekves, 2019).

Acknowledgments

We thank Valeria Pfeifer for her assistance with the stimuli creation, Thijs Rinsma and Maarten van den Heuvel for their help with the implementation of the pre-tests and support setting up the studies, and Ksenija Slivac for advice regarding the analysis of the data.

5. Appendix

List of sentence stimuli and associated objects for Experiment 2.

	Subject	Perfective	Progressive		Object SC/NSC	Object U
1	Daniel Bonny	bit	was biting	the the	sandwich. cookie.	scarf. backpack.
2	Robert Linda	boiled	was boiling	the the	pasta. water.	paper plane. necklace.
3	Noah Daisy	burned	was burning	the the	toast. match.	Coca Cola. smoothie.
4	Jacob Charlotte	chopped	was chopping	the the	onion. leek.	spoon. shoes.
5	Matt Olivia	cubed	was cubing	the the	cucumber. cheese.	nail. package.
6	Ben Ella	diced	was dicing	the the	tomatoes. garlic.	doll. boomerang.
7	Alex Taylor	dipped	was dipping	the the	crackers. bread.	Lego tower. flower.
8	Tony Sue	fried	was frying	the the	mushrooms. steak.	frisbee. puzzle.
9	John Patty	grated	was grating	the the	parmesan. carrot.	chips. cushion.
10	Michael Liz	juiced	was juicing	the the	grapefruits. oranges.	postcard. pencil.
11	Will Jenny	mashed	was mashing	the the	peas. beans.	keys. card house.
12	David Karen	melted	was melting	the the	ice cream. chocolate.	shopping list. clothes.
13	Richard Helen	peeled	was peeling	the the	banana. mandarin.	handbag. magazine.
14	George's nephew Sarah's wife	roasted	was roasting	the the	chicken. peppers.	fork. marbles.
15	Anthony's brother Julia's sister	skinned	was skinning	the the	potato. apple.	hat. waffle.
16	Ryan's mum Hazel's friend	sliced	was slicing	the the	bagel. pizza.	napkin. bracelet.
17	Henry's sibling Jade's twin	squeezed	was squeezing	the the	limes. lemons.	cactus. pot.
18	Brian's son Steve's daughter	broke	was breaking	the the	vase. clock.	crossword. soup.
19	Lucy's husband Susan's niece	colored	was coloring	the the	drawing. painting.	wallet knot.
20	Evan's fiancé Ruby's mother	lit	was lighting	the the	cigarette. candle.	candy. rock.
21	Edward's cousin Chloe's aunt	scratched	was scratching	the the	screen. phone.	salad. newspaper.

22	Emily's uncle Tim's dad	spilled	was spilling	the the	paint. whiskey.	plastic bag. lamp.
23	Nathan's grandad Claire's grandma	stained	was staining	the the	bib. shirt.	balloon. juice.
24	Laura's colleague Rachel's co-worker	shredded	was shredding	the the	contract. bank statements.	pear. ring.

CHAPTER 4

CAN GRAMMATICAL ASPECT MODULATE EVENT MEMORY?[8]

Abstract

The present study examined the effects of aspect on event memory using a web-based setup. Employing two experiments allowed us to systematically assess the role aspect may play in event memory, as it can be a cue prior to perception (Experiment 1) and/or a means for memory encoding after event apprehension (Experiment 2). In Experiment 1, for each trial, participants had to read aloud a sentence in which aspect was manipulated (e.g., "*She chopped/was chopping an onion.*"), and thereafter they watched video clips of the described event. Videos either stopped playing before the end of the event (in-progress) or at completion (ceased). In addition, event type was manipulated; in half of the events object state changed (SC, e.g., to chop an onion) and in the other half object state did not change (NSC, e.g., to type on a keyboard). Experiment 2 used the same stimuli, but reversed the order such that participants first watched the video and then read the sentence. In a surprise recognition memory task, participants saw video stills and had to indicate whether each still correctly showed how the video had ended. We analysed memory accuracy, RTs on the correct responses, and signal detection measures d' and C, reflecting sensitivity and response bias, respectively. The results of both experiments highlight an interaction of Event Type and Video Status; participants were more accurate, faster, and more sensitive in responding to in-progress videos of NSC events (as compared to in-progress SC videos), and in responding to ceased videos of SC events (as compared to ceased NSC videos). In Experiment 2, we additionally observed an interaction of Aspect and Video Status, such that memory for ceased videos was better for events described in the progressive (compared to the perfective) and better for in-progress videos described in the perfective (compared to the progressive). The different patterns of results across the experiments are discussed in terms of whether language was introduced as a tool for memory before or after viewing the event.

[8]Manuscript in preparation: Misersky, J., Kamentski, A., & Flecken, M. (in prep). Can grammatical aspect modulate event memory?

1. Introduction

Events are at the core of human cognition, and the way we experience, understand and memorise events occurs at the interface of vision, action, and language. As a result, the field of event cognition has garnered more attention in research in cognitive science and beyond. Broadly speaking, an event is framed by the location and time in which it occurs. In addition to these framing pillars, the specifics of the event also include event participants (e.g., people and objects involved in the event), and how they relate to one another. When experiencing something, we use so-called working models, which support us in mapping out how the experience we are having might unfold. A working model takes into account our prior experiences, as well as the currently perceived dimensions of time, space and event participants, and we use this to 'chunk' the ongoing activity around us into discrete events. When it becomes too difficult to accurately predict how an event will continue, an event boundary is detected, the experience is segmented and subsequently conceptualised as an event with beginning and end (Event Segmentation Theory; EST, Zacks & Swallow, 2007). Accurate prediction becomes challenging when there is a change along one (or more) of the aforementioned event dimension(s), for example when an object undergoes a change over time. When a boundary is encountered, the event segments (or 'chunks') are stored as memory representations called event models (Radvansky & Zacks, 2014), which can be employed again when making sense of new experiences. Over time, existing event models become adapted to represent variations of comparable events in memory (cf., Event Horizon Model, Radvansky & Zacks, 2014).

While EST is mostly concerned with the detection of event boundaries, more recently, Altmann and Ekves (2019) have put forward a theoretical account of how we conceptualise what happens between the event boundaries. At the core of the internal event structure, they envision the tracking of objects throughout an event (Intersecting Object Histories; IOH). Both accounts describe related facets of event cognition, and highlight the importance of change in the conceptualisation of events. In EST, change leads to the detection of event boundaries, and specifically, the end of the event (i.e., the event's right boundary). IOH specifies that objects change in state, with the beginning, intermediate and end states tracked during event conceptualisation. Empirical work suggests multiple object states may be active simultaneously and even compete (Hindy, Altmann, Kalenik, & Thompson-Schill, 2012; Solomon, Hindy, Altmann, & Thompson-Schill, 2015). In a sense, the beginning and end states of the object characterise the event boundaries. In addition, the tracking of intermediate object states brings with it the understanding of changes along other event dimensions including time, location, but also changes in other event participants such as people or additional objects. Accurate tracking of object states along the whole trajectory of change will allow for successful prediction of how an experience will continue. Tracking thus supports event segmentation, and by extension the storage of event models in memory.

In sum, our understanding of events is guided by our prior experience with similar events. Tracking change along event dimensions such as object change over time has been discussed at the core of event conceptualisation. As described above, events can be experienced directly. Importantly, though, we can also experience events indirectly, through language or linguistic input. Through descriptions of events, for example in the news or a story told by a friend, language may influence the creation of a working model, and how this information is then used to understand new events, and how an event is stored in memory.

Indeed, behavioural research on eyewitness testimonies has shown that the specific wording of event descriptions affected which details of an event were remembered (Loftus & Palmer, 1974). Participants watched video clips of car accidents, and thereafter they were asked to estimate the speed of the cars. The leading questions varied in wording ("*How fast were the cars going when they smashed into/collided with/bumped into/contacted/hit each other?*"), and speed estimates were highest for participants having read "*smashed into*". A week later, those same participants also reported having seen broken glass, which was not present in any of the video clips. Despite all participants seeing the same video clips, the labelling description of the videos lead to the events being stored in memory differently across participants. In a recent study using a similar format, Goldschmied, Sheptock, Kim and Galily (2017) aimed to replicate these findings using sports events (hockey games) as thematic backdrop. Similar to Loftus and Palmer's (1974) study, participants watched a collision, and were asked to estimate the speed of the hockey players involved. The results of Palmer and Loftus (1974) did not replicate to sports events when participants were presented with the leading questions after viewing the video. However, when they watched the video with running commentary, they found that speed estimates differed depending on which verb was used in the commentary. The results of both studies indicate that the meaning of single words can quickly guide our attention and conceptualisation of incoming perceptual input, as well as how we store it in memory. Though future research seems needed, it is also clear that these framing effects on memory may differ depending on the types of events being perceived as well as when the linguistic input is provided (namely prior, during or after event perception). Acting as labels, single word cues also have the power to activate representations prior to actually (visually) experiencing a percept (Lupyan, 2012; Lupyan, Abdel Rahman, Boroditsky, & Clark, 2020). For example, upon hearing or reading the word "*onion*", the representation of an onion is activated prior to seeing it. In addition, a label allows for quicker perception of an object, such that when the image of, for example, an onion is degraded or masked, it is recognised quicker when a label is provided prior to visual stimulation. In their Experiment 1, Kang, Eerland, Joergensen, Zwaan and Altmann (2019) provided participants with object nouns, and then showed them pictures of objects, asking them to indicate whether the object shown in the picture was the same as described by the object noun. The results indicate that providing a single label such as "*onion*" leads to participants responding slower to the image of an onion having undergone change (e.g., a chopped onion) compared to when the onion was in its original, unchanged state. In sum, single words can allow for the quick apprehension of (prototypical) objects, but may also lead to false memories.

Beyond these single labels, effects of language on event cognition can take effect at the sentence-level (e.g., Sauppe & Flecken, 2021 on rapid scene apprehension), and also through specific grammatical features. For example, in Experiments 2 and 3, and based on Zwaan and Pecher (2012), Kang et al. (2019) embedded object labels in sentences. Adding a sentence context led to a different pattern of results compared to what they saw in Experiment 1: When the semantics of the verb suggested a change in object state (e.g., drop an ice cream), participants responded faster to the picture showing the object having undergone a change compared to the object in its original state. This highlights the importance of the sentence in the construction of an event model, activating object properties corresponding to the verb semantics. Recently, verb semantics have been found to guide the understanding of change-of-state events also in incomplete sentences only consisting of subject and verb yet omitting the object noun (see Chapter 3, Experiment 2). This indicates that people are able to represent object change even when an object is unknown. In addition to verb semantics, perfective aspect—a grammatical means of expressing event time as ceased—draws the attention to the goal state of an event, foregrounding the changed state of an object (Misersky, Slivac, Hagoort, & Flecken, 2021). By contrast, progressive aspect may diffuse event boundaries (Madden & Zwaan, 2003; Madden-Lombardi, Dominey, & Ventre-Dominey, 2017), including the end state of the event when verb semantics suggest an event with inherent endpoint (cf. *telicity*, Comrie, 1976; Misersky et al., 2021).

Turning to event memory specifically, Sakarias and Flecken (2019), who tested Dutch and Estonian native speakers on their memory for events with and without state change, also reported saliency of object states in event conceptualisation and memory. Unlike Dutch, Estonian uses case-marking on the noun to express changes in the state of objects (partial or full change). Participants watched video clips, which ended in the action ceased or still ongoing. In the verbal condition, they were asked to describe what had happened in the video. At the end of the study, participants also engaged in a surprise memory task, in which they had to choose the correct still depicting how the video they had seen ended (two-alternative forced choice task, 2AFC; with one video still showing the ceased and the other the ongoing version). Case-marking boosted participants' performance, such that Estonian speakers were better able to accurately recognise the event endings compared to the Dutch speakers. Cross-linguistic differences in grammatical means to convey information on object states (here: through case-marking) thus influences how events are encoded, at least under verbal conditions. In addition, participants generally allocated more attention to objects undergoing a change-of-state. Recently, Santín, van Hout and Flecken (2021) also took a cross-linguistic approach to study event memory. Similar to Sakarias and Flecken (2019), the authors had participants watch video clips of short action events, and then tested their subsequent memory for video endings in a surprise memory task. Specifically, the authors looked at event dynamics (ceased vs ongoing action) and event type (state change vs no state change) under verbal and non-verbal conditions. The experiment was carried out with Spanish and Mandarin Chinese speakers, who differ in the way they express culmination (i.e., the endpoint or -state of an event after which the action can no longer continue) of events at the verb level. Across both language groups, an interaction effect indicated increased memory performance in relation to ceased events depicting a change-of-state. The effect was stronger in the verbal condition, where participants produced a description of the event in their native language, suggesting language aided memory performance.

Similar results suggesting increased performance under conditions where language can be used as a tool have also been observed for motion events (e.g., Skordos, Bunger, Richards, Selimis, Trueswell & Papafragou, 2020; Trueswell & Papafragou, 2010). Most recently, Skordos et al. (2020) have specifically manipulated whether, when and how language was presented. In a cross-linguistic comparison, they tested native speakers of English and Greek, who differ in their preference to encode manner and path in motion verbs which may bias encoding and memory of motion events. In Experiment 1, participants watched animations of motion events silently and were told that their memory for the events would be tested. In Experiment 2, they watched the animations and had to produce a descriptive verb after the event had finished. In Experiment 3, the participants were provided descriptive verbs prior to viewing the animations. In Experiments 2 and 3, participants were unaware that their memory for the clips (i.e., deciding whether clips where the same or different to previously watched ones) would be tested. There were no cross-linguistic differences between the groups in memory accuracy. However, regardless of language background, the authors did note effects of language: When language, specifically path verbs, were either being produced by the participants (Experiment 2) or provided by the experimenter (Experiment 3), memory for manner of motion in the events was reduced. Skordos et al. (2020) conclude that language can offer tools to categorise event information. In their specific study language impaired memory, though they note that language may also boost memory in other event contexts.

In sum, the literature on event cognition points towards language supporting memory for information relevant to (change-of-state) events, particularly when language is being employed during the task (e.g., Santín et al., 2021; Sakarias & Flecken, 2019). In addition, the abovementioned studies are in line with accounts of a general cognitive bias for goal states (Lakusta & Carey, 2015; Altmann & Ekves, 2019), suggesting heightened sensitivity to event goals and endpoints. The goal bias was also observed in a study looking at event memory, in which language was not explicitly manipulated (Strickland & Keil, 2011). The authors presented participants with video clips of launching events (e.g., kicking, throwing), in which the specific moment of contact or release was omitted. People reported having seen the moment of contact even when this was not the case, which the authors take as evidence of a tendency to make inferences about causality in goal-oriented events. Causality and causal relations between, for example, actors and objects, have been discussed as central to event memory organisation (Zacks, 2020).

While language can be used as a tool for memory performance, it is important to note that memory can be guided by linguistic input in two ways (cf. Skordos et al., 2020). Firstly, by cueing prior to experiencing the event, where language can guide the observer's attention to specific event features (language for perception). Secondly, language can have an effect on how an event is remembered by marking event features (language for encoding). A general cognitive bias for goal states may affect our memory of change-of-state events additionally. Zacks suggests time may be a 'potential organizer of events in memory' (2020, p. 174), and a goal bias may thus interact with linguistic markers of time, such as grammatical aspect. Indeed, aspect has been shown to add subtle nuances to understanding the trajectory of changing states (Misersky et al., 2021; Chapters 2 and 3), though the effect of aspect on event memory requires further study.

1. 1. The Present Study

We were interested in studying the role of grammatical aspect on event memory, to assess whether language describing such events, and specifically grammar (here: aspect signalling a ceased or ongoing event in the past), would affect how events are remembered. Thus, in both experiments, participants were given language as a tool to shape event memory, yet they were unaware that memory would be tested (cf. Skordos et al., 2020). We aimed to take a more comprehensive view on language as a guide to memory by looking at language prior to and after event experience. Providing language prior to the event could create expectations regarding the input and may additionally bias the processing and representation of the input. To this end, in Experiment 1, we looked at language for perception by employing a linguistic cueing paradigm in which participants read descriptive sentences aloud before watching video clips of events. In Experiment 2, we were interested in language for encoding, as this has been suggested to affect the memory of experienced events (Skordos et al., 2020). In Experiment 2, participants first watched video clips and then were provided with descriptive sentences to read aloud. Similar to previous work, we opted to focus on events where objects either undergo a state change (e.g., to peel an apple, SC) or no state change (e.g., to clap hands, NSC). In both experiments the sentences followed SVO structure, and the verb aspect was manipulated (e.g., "*She chopped/was chopping an onion.*"). In addition, we manipulated the video status of the event, such that half of the clips ended with a ceased action (the actor withdrew the hands from the object after completing the action) or an in-progress action (the actor still had the hands on the object whilst carrying out the action). Crucially, we engaged participants in a surprise memory task after watching the events, tapping into their memory for how the video clips ended.

In line with previous studies, and in line with a general bias for goals (Lakusta & Carey, 2015), we expected SC events to be remembered better overall. We also expected Event Type (SC vs NSC) to interact with Video Status, such that when the video showed a ceased state change event (ceased—SC) this should be remembered better than a ceased no state change event (ceased—NSC), based on the results of Santín et al. (2021). Crucially, we were interested in how Aspect would interact with Video Status: Previous work has shown that perfective aspect can cue for an event's endpoint (e.g., Misersky et al., 2021), including the object's complete state change. The perfective is thus more likely to constrain the event model to holistic represenions compared to the progressive, leading to overall processing of ceased events being facilitated. Constraining of the event model through the perfective may facilitate storage of an event model in memory, specifically if the event model suggested by the sentence matches what participants see. Thus, a match between Aspect and Video Status could improve event memory. Alternatively, a mismatch between Aspect and Video Status could also enhance memory: When aspect focuses attention on specific features relevant to the event's ending, which are then not observed (Experiment 1) or when it describes an event model different to what has been witnessed (Experiment 2), this could lead to increased conservativeness regarding what is committed to memory. The effects of Aspect and/or Video Status could also be modulated by event semantics. Going back to the notion of change in events, SC events code for the degree of change, as do Aspect and Video Status. Since SC events encapsulate change by definition, effects of (in)congruence or between Aspect and Video Status may be enhanced for this Event Type. In line with this, NSC events lack change and clear goals, meaning the effect between Aspect and Video Status may be less pronounced for this Event Type.

2. Pre-Test

In light of a larger cross-linguistic collaborative project, the pre-test was carried out in four languages: English, German, Dutch, and Russian. After careful consideration, items were selected in consensus with the pre-test data from all languages. For the main study 48 English items were selected, and the results of the English pre-test are reported in line with the focus of this paper.

Participants

Thirty native speakers of English (10 female, 2 non-binary, 18 male, 18-45 years of age, *M* = 25.22, *SD* = 7.22) participated in this pre-test. The pre-test 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 monetary compensation for their time (6 GBP per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, The Netherlands.

Materials and Design

For the materials, we partly used video clips (*N* = 27) from Santín et al. (2021). Further, we filmed 82 additional videos. All videos included female actors in black shirts sitting at a table and performing various actions. Of the resulting videos, 58 presented state change events where an object underwent a visible change, and 51 showed events where no state change was observed. Videos were cut such that the actors' faces were out of the frame, meaning the focus was on their hands and the performed action. Further, for each unique clip, there were an in-progress version of the event (video stopped at 75% played), and a ceased version of the event (video stopped at 100% played). Since the length of the clips was shorter for in-progress clips compared to the ceased clips, speed was adjusted so that versions had comparable length[9]. In addition, there were descriptions accompanying the video clips, framing the events either as a change-of-state event (e.g., to peel an apple) or a no-change-of-state event (e.g., to clap hands).

Procedure

Participants gave informed consent and then answered a set of background questions (age, gender, nationality, additional languages). They then watched video clips and answered three sets of questions. First (Question 1), they had to rate on a 7-point scale whether the action was still ongoing by the end of the video (1 = action no longer ongoing; 7 = action still ongoing). This was to test whether the two versions of each video were perceived as in-progress and ceased. Second (Question 2), they were provided with a description (e.g., "*to peel an apple*") and were asked to rate on a 7-point scale whether the description fits with the video they saw (1 = does not fit at all; 7 = fits very well). In cases where the rating was 3 or lower, a text box appeared asking participants to provide a description they deemed more fitting. This was to ensure that our chosen descriptions, which were to be used in the main experiment, were perceived as descriptive of the video clips. Lastly (Question 3), for the ceased videos, participants were asked to rate on a 7-point scale how much the object in the video underwent a change (1 = no change at all; 7 = object is changed completely). This was to ensure participants perceived the event types displayed in the videos as state change (SC) and no state change (NSC) events. Participants saw a total of 103 trials.

Data Analysis and Results

For Question 1 ("*To what extent was the action still ongoing at the end of the video?*"; 1 = no longer ongoing, 7 = still ongoing), mean ratings for in-progress versions were M = 6.61 (SD = 0.93), and for ceased videos M = 1.93 (SD = 1.67). For Question 2 ("*How well does this description fit the event in the video?*"; 1 = does not fit at all, 7 = fits very well), the mean rating was M = 6.66 (SD = 0.70). When giving ratings of 3 or below, participants were prompted to provide a description themselves. This, however, was rarely the case with descriptions receiving high ratings in the majority of trials (range 99-103 trials, out of 103). For Question 3 ("*To what degree did the object change from its initial state (the way it looked at the beginning of the video)?*"; 1 = no change at all, 7 = the object is changed completely), mean ratings for NSC were M = 2.43 (SD = 1.81), and for SC videos M = 5.18 (SD = 1.54). Outlier rating scores for all questions were established using z-scores (MAD method), with a cut-off of 2.5 (Leys et al., 2013). For the main experiment, 48 unique events (24 SC, 24 NSC) and associated descriptions were chosen (see Appendix for the event descriptions).

[9]For in-progress videos, video durations ranged from 2.82 seconds to 5.88 seconds (M = 4.95, SD = 0.61). For ceased videos, video durations ranged from 4.31 seconds to 6.04 seconds (M = 5.03, SD = 0.35). Importantly, all the videos were edited such that the in-progress and ceased versions of the same event were as close as possible in duration. Differences in duration of in-progress and ceased videos were not statistically significant ($t(75.527) = -0.733$, $p = 0.466$).

3. Experiment 1: Language for Perception (Sentence First)

3. 1. Methods

Participants

Fifty-six native speakers of English (40 female, 16 male, 18-44 years of age, *M* = 29.84, *SD* = 7.19) participated in this study. The experiment 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 monetary compensation for their time (6 GBP per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, The Netherlands.

Materials and Design

For the materials, and based on the pre-test, we used 48 unique event video clips (24 SC, 24 NSC) and associated descriptions. Eight unique lists were created to account for all combinations of Event Type (SC vs NSC), Aspect (perfective vs progressive), Video Status (in-progress vs ceased), and correct answer in the probe (Yes vs No). Items within each list were presented randomly per participant, yet the order of items in the encoding phase was the same as the order of probes displayed in the memory test.

Procedure

Participants gave informed consent, and then answered a number of background questions (age, gender, nationality, additional languages). The study was split into three blocks: The Encoding block, the Corsi block, and the surprise Recognition memory block (see Fig. 1). They received the instructions for the Encoding block, as well as a try-out for their microphone and recording. In the Encoding block, for each trial, they saw a sentence on screen (SVO; e.g., "*She chopped/was chopping an onion.")* and were asked to read it aloud for recording. When done, they could stop and submit the recording. They then advanced to the video, which was only played once, then disappeared from the screen. Participants were then able to press Continue to advance to the next trial. After seeing all trials, instructions for the Corsi block appeared. In the Corsi blocks tapping task, participants saw outlines of nine squares arranged on the screen. With each trial, a set of the squares would flash in black in a particular order, and participants had to indicate the correct order by clicking on the squares. Participants received two practice trials (including being shown the correct expected response) before doing the task. The length of the Corsi block was dependent on participants' performance on the task with each trial becoming more difficult (i.e., more blocks being flashed thereby requiring increased working memory performance to retain the correct order) than the previous one. The task ended when participants gave incorrect responses in two consecutive trials. In the surprise Recognition memory block, participants were shown stills of the last frame of the video clips, either in the in-progress or ceased version, and asked to indicate whether this was how the video ended. They had to respond with Yes or No, and were asked to do this as quickly as possible. There was a timeout after 5000 milliseconds and a message "*Too slow!*" appeared. Advancing to the next item was then possible via button press.

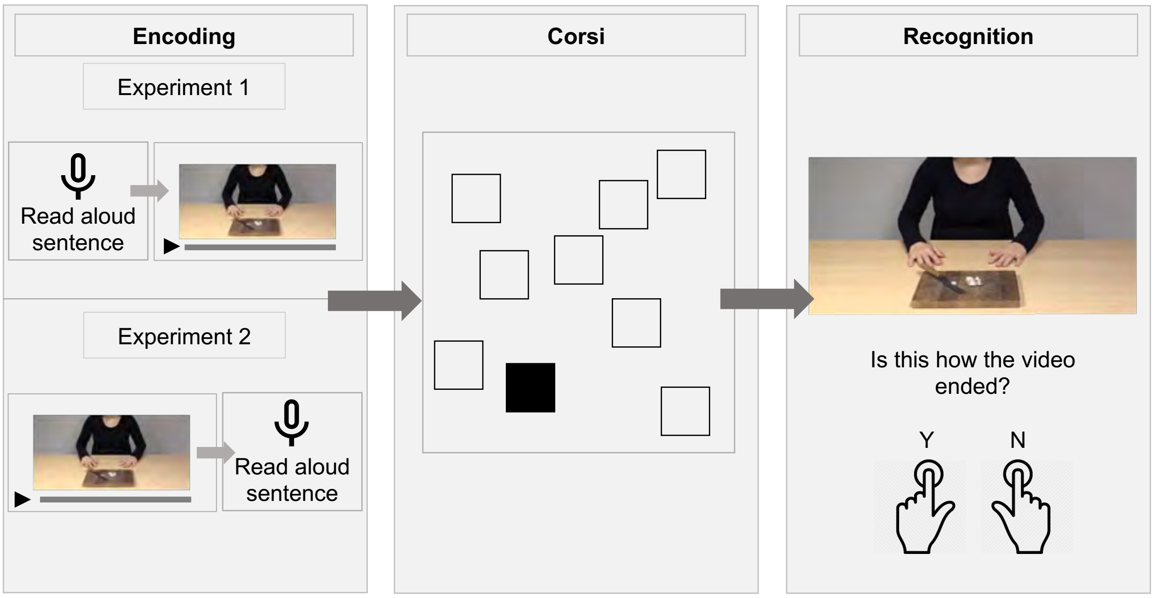


Figure 1: *Example of the study setup.*

Data Analysis

For the analysis, we focused on four measures from the Recognition memory block: First, accuracy on the responses to the video stills. Second, log-transformed reaction times (RTs) associated with correct responses to the video stills. Third, we did signal detection analyses focused on participants' d-prime (d') and C scores reflecting sensitivity and individual response bias, respectively. Scores for the signal detection analyses were calculated in R using the d-prime function from the psycho package (Makowski, 2018).

We analysed sum-coded data using linear effects modelling in R (Version 3. 6. 3). All models were fitted using the packages lme4 (Version 1.1-26, Bates, Maechler, Bolker, & Walker, 2015) or stats (R Core Team, 2013). Based on our manipulation and research questions, we used as fixed factors Event Type (SC vs NSC), Aspect (perfective vs progressive), and Video Status (in-progress vs ceased), and their interactions. In addition, we included the individual scores from the Corsi block in the models as a fixed effect (control variable). We opted to use the same model structure for accuracy and RTs, and for d' and C.

For the random effects (RE) structure, we started out with intercept-only effects, including both participant and verb for the measures of accuracy and RTs, and participant for d' and C. Goodness-of-fit of the resulting models was then assessed using the rePCA function (lme4, Bates et al., 2015), and the RE structure was adjusted and re-assessed according to this same method until model fit was optimised. For the d' and C analyses, and due to singular fit of the d' model, a fixed-effects-only model was fitted. For the respective models, please refer to the Results section below.

3. 2. Results

Accuracy. The main effect of Event Type approached significance (p = 0.0596), suggesting that overall SC events were remembered better (M = 0.564, SD = 0.496) compared to NSC events (M = 0.527, SD = 0.449). As can be seen in Fig. 2, there was also an interaction of Video

Status and Event Type ($p < .001$); SC events were remembered better in the ceased version ($M = 0.694$, $SD = 0.461$) compared to the in-progress version ($M = 0.433$, $SD = 0.496$). In contrast, NSC events were remembered better in the in-progress version ($M = 0.626$, $SD = 0.4841$) as compared to the ceased version ($M = 0.428$, $SD = 0.495$).

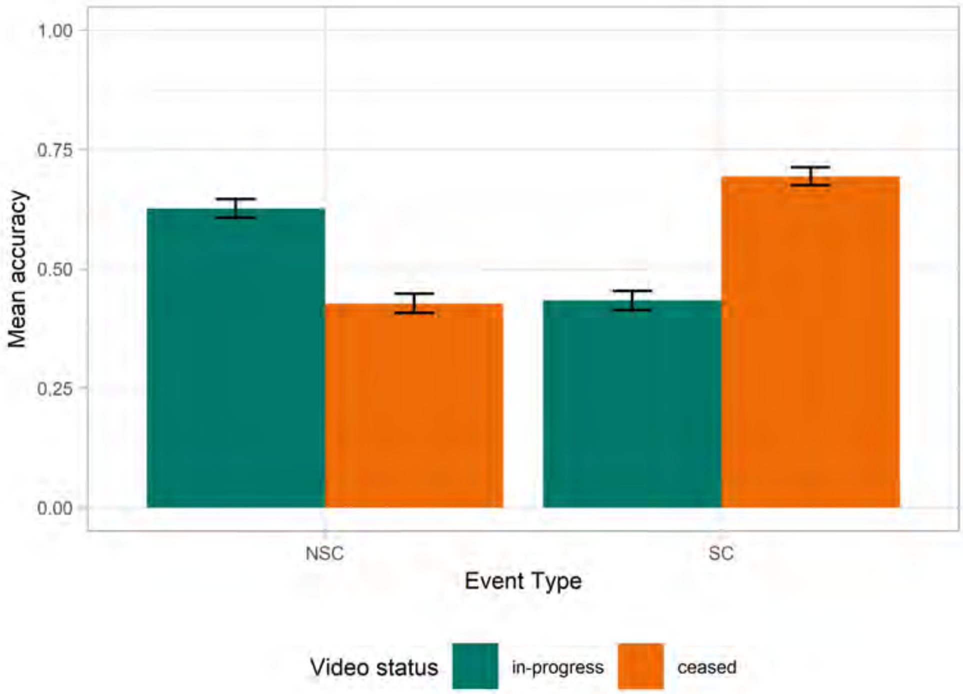


Figure 2: *Mean accuracy (in %) for Event Type (NSC vs SC) and Video Status (in-progress vs ceased) ± 1 SEM.*

There were no (interaction) effects of Aspect. A significant main effect of Corsi score ($p = .030$) suggests a better working memory score led to better accuracy performance.

Table 1. Accuracy analysis: Output of the model on responses to the video stills (glmer (Accuracy ~ Event Type * Video Status * Aspect + Corsi score + (Video Status | Participant) + (Video Status | Verb))

Fixed Effects	*β* Estimate	SE	*z*-Value
Intercept	-0.259	0.226	-1.143
Event Type	-0.088	0.047	-1.884
Video Status	-0.073	0.129	-0.563
Aspect	0.014	0.045	0.319
Corsi score	0.059	0.027	2.165*
Event Type : Video Status	0.548	0.067	8.160***
Event Type : Aspect	0.009	0.045	0.205
Video Status : Aspect	-0.055	0.045	-1.235
Event Type : Video Status : Aspect	0.023	0.045	0.507

***$p < 0.001$; **$p < 0.001$; *$p < 0.05$

Reaction times. No main effects were observed for the RTs of the correct responses. There was a significant interaction of Video Status and Event Type ($p < .001$; Fig. 3); SC events were responded to faster in the ceased version (*M* = 2081.824, *SD* = 895.638) compared to the in-progress version (*M* = 2294.717, *SD* = 960.992).

In contrast, NSC events were responded to faster in the in-progress version (*M* = 2132.358, *SD* = 904.716) as compared to the ceased version (*M* = 2266.755, *SD* = 904.86). There were no (interaction) effects of Aspect.

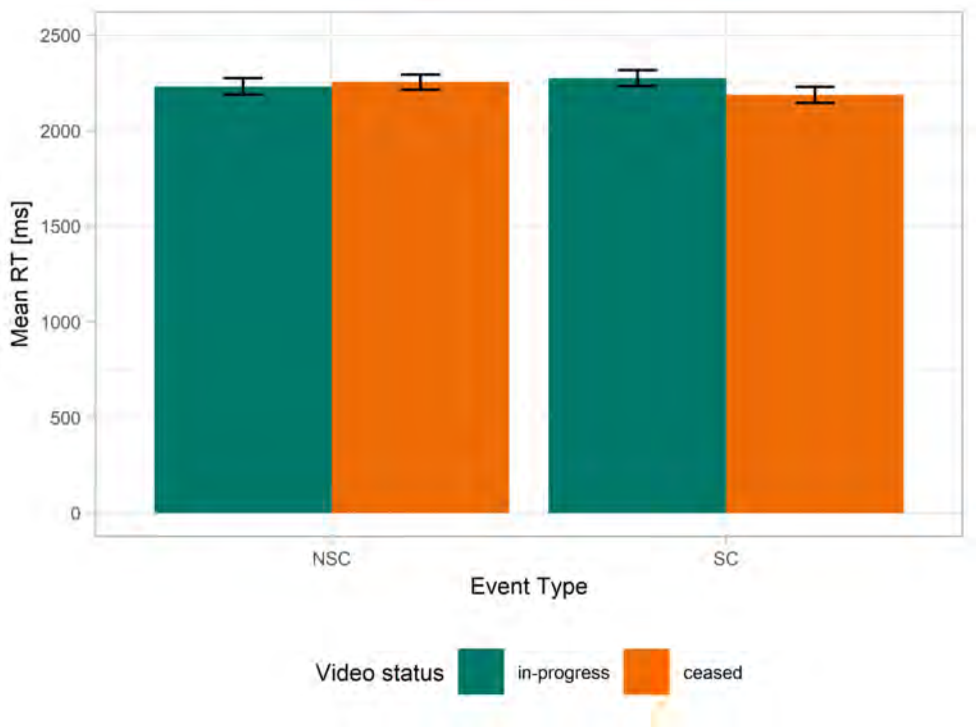


Figure 3: *Mean RTs (in milliseconds) for Event Type (NSC vs SC) and Video Status (in-progress vs ceased) ± 1 SEM.*

Table 2. RT analysis: Output of the model on log RTs of the correct responses to the video stills (lmer (logRT ~ Event Type * Video Status * Aspect + Corsi score + (Video Status | Participant) + (Video Status | Verb))

Fixed Effects	*β* Estimate	SE	*t*-Value
Intercept	7.756	0.184	42.078***
Event Type	0.009	0.01	0.883
Video Status	0.005	0.01	0.481
Aspect	-0.0003	0.008	-0.041
Corsi score	-0.018	0.022	-0.819
Event Type : Video Status	-0.042	0.009	-4.750***
Event Type : Aspect	0.007	0.008	0.893
Video Status : Aspect	0.0003	0.008	0.031
Event Type : Video Status : Aspect	-0.014	0.008	-1.629

****p* < 0.001; ***p* < 0.001; **p* < 0.05

Signal detection. The d' analysis was carried out on the z-scores of hits minus false alarms. No main effects were observed for sensitivity (d').

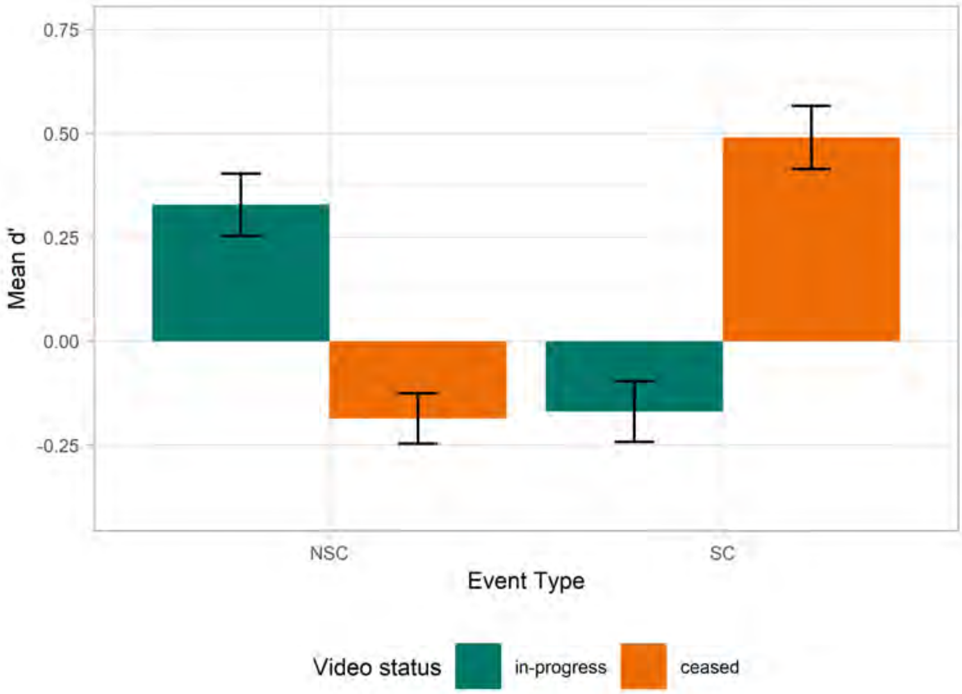


Figure 4: *Mean d' scores for Event Type (SC vs NSC) and Video Status (in-progress vs ceased) ± 1 SEM.*

There was a significant interaction of Video Status and Event Type ($p < .001$, Fig. 4); SC events had higher d' scores, meaning they received more hits than false alarms in the ceased version ($M = 0.490$, $SD = 0.772$) compared to the in-progress version ($M = -0.169$, $SD = 0.744$). By contrast, NSC events had lower d' scores, which showed in more hits than false alarms in the in-progress version ($M = 0.329$, $SD = 0.763$) as compared to the ceased version ($M = -0.186$, $SD = 0.613$). There were no (interaction) effects of Aspect.

Table 3. d-prime analysis: Output of the model on sensitivity (lm (d-prime ~ Event Type * Video Status * Aspect + Corsi score)

Fixed Effects	*β* Estimate	SE	*t*-Value
Intercept	-0.129	0.16	-0.809
Event Type	-0.045	0.036	-1.252
Video Status	-0.036	0.036	-1.017
Aspect	0.012	0.036	0.334
Corsi score	0.031	0.019	1.575
Event Type : Video Status	0.293	0.036	8.231***
Event Type : Aspect	0.011	0.036	0.322
Video Status : Aspect	-0.032	0.036	-0.907
Event Type : Video Status : Aspect	0.012	0.036	0.349

***$p < 0.001$; **$p < 0.001$; *$p < 0.05$

In addition, we analysed C (i.e., the number of standard deviations from the midpoint between the signal and signal-to-noise distribution), an index for response bias. The main effect of Event Type was approaching significance (p = .051), suggesting participants were less conservative in their responses to SC events (M = -0.114, SD = 0.405) compared to NSC events (M = -0.04, SD = 0.368). There were no interaction effects.

Table 4. Criterion analysis: Output of the model on bias (lm (c ~ Event Type * Video Status * Aspect + Corsi score)

Fixed Effects	β Estimate	SE	t-Value
Intercept	-0.032	0.086	-0.378
Event Type	0.037	0.0191	1.9657
Video Status	-0.022	0.0191	-1.179
Aspect	-0.005	0.0191	-0.25
Corsi score	-0.006	0.01	-0.536
Event Type : Video Status	0.003	0.0191	0.135
Event Type : Aspect	-0.002	0.0191	-0.103
Video Status : Aspect	-0.011	0.0191	-0.573
Event Type : Video Status : Aspect	-0.01	0.0191	-0.527

***p < 0.001; **p < 0.001; *p < 0.05

4. Experiment 2: Language for Encoding (Video First)

4. 1. Methods

Participants

Fifty-six native speakers of English (42 female, 14 male, 18-43 years of age, M = 28.45, SD = 7.39), participated in this study. The experiment 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 monetary compensation for their time (6 GBP per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, The Netherlands.

Materials and Design

The materials and lists were identical to Experiment 1.

Procedure

The procedure was identical to Experiment 1 with the exception that for each trial in the Encoding block, the video was played first and then the accompanying sentence was read.

Data Analysis

The steps taken for data analysis were identical to Experiment 1. For the respective models, please refer to the Results section below.

4. 2. Results

Accuracy. We observed no main effects. There was an interaction of Video Status and Event Type (*p* < .001; Fig. 5 left); like in Experiment 1, SC events were remembered better in the ceased version (*M* = 0.737, *SD* = 0.44) compared to the in-progress version (*M* = 0.406, *SD* = 0.491). In contrast, NSC events were remembered better in the in-progress version (*M* = 0.686, *SD* = 0.464) as compared to the ceased version (*M* =0.400, *SD* = 0.490).

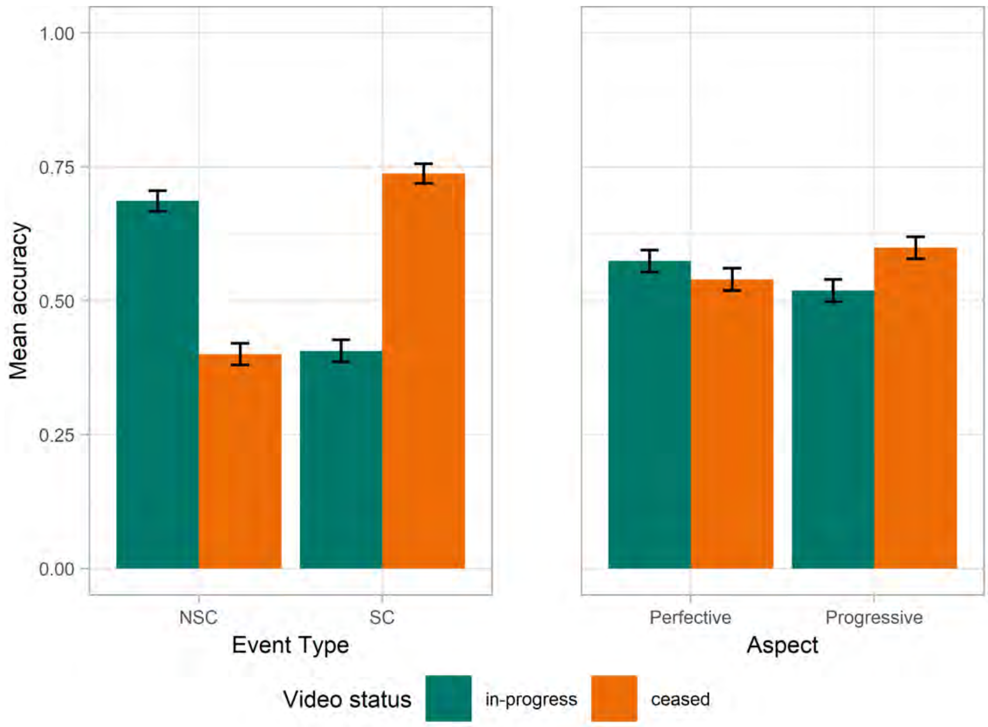


Figure 5: *Left: Mean accuracy (in %) for Event Type (SC vs NSC) and Video Status (in-progress vs ceased) ± 1 SEM. Right: Mean accuracy (in %) for Aspect (perfective vs progressive) and Video Status (in-progress vs ceased) ± 1 SEM.*

In addition, we observed an interaction of Video Status and Aspect (*p* = .004, Fig. 5 right). Participants' memory was better for ceased videos when sentences were in the progressive (*M* = 0.599, *SD* = 0.491) compared to the perfective (*M* = 0.537, *SD* = 0.499). The opposite was true for in-progress videos, which were remembered better with perfective (*M* = 0.574, *SD* = 0.495) compared to progressive sentences (*M* = 0.519, *SD* = 0.500). The main effect of Corsi score approaching significance (*p* = .05) suggests a better working memory score led to better accuracy performance.

Table 5. Accuracy analysis: Output of the model on responses to the video stills (glmer (Accuracy ~ Event Type * Video Status * Aspect + Corsi score + (Video Status | Participant) + (1 | Verb))

Fixed Effects	*β* Estimate	SE	*z*-Value
Intercept	-0.162	0.23	-0.704
Event Type	-0.058	0.047	-1.24
Video Status	-0.054	0.093	-0.58
Aspect	-0.006	0.046	-0.133
Corsi score	0.058	0.03	1.959
Event Type : Video Status	0.704	0.047	14.951***
Event Type : Aspect	-0.024	0.046	-0.522
Video Status : Aspect	0.132	0.046	2.868**
Event Type : Video Status : Aspect	-0.002	0.046	-0.046

***$p < 0.001$; **$p < 0.001$; *$p < 0.05$

Reaction times. For the RTs of the correct responses, we observed a main effect of Event Type ($p = 0.014$), meaning that overall SC events were responded to faster (M = 2139.193, SD = 745.098) compared to NSC events (M = 2251.015, SD = 772.700). In addition, there was a significant interaction of Video Status and Event Type ($p < .001$; Fig. 6); SC events were responded to faster in the ceased version (M = 2096.055, SD = 723.204) compared to the in-progress version (M = 2219.343, SD = 779.085). In contrast, NSC events were responded to faster in the in-progress version (M = 2196.701, SD = 782.430) as compared to the ceased version (M = 2345.941, SD = 747.692). There were no (interaction) effects of Aspect.

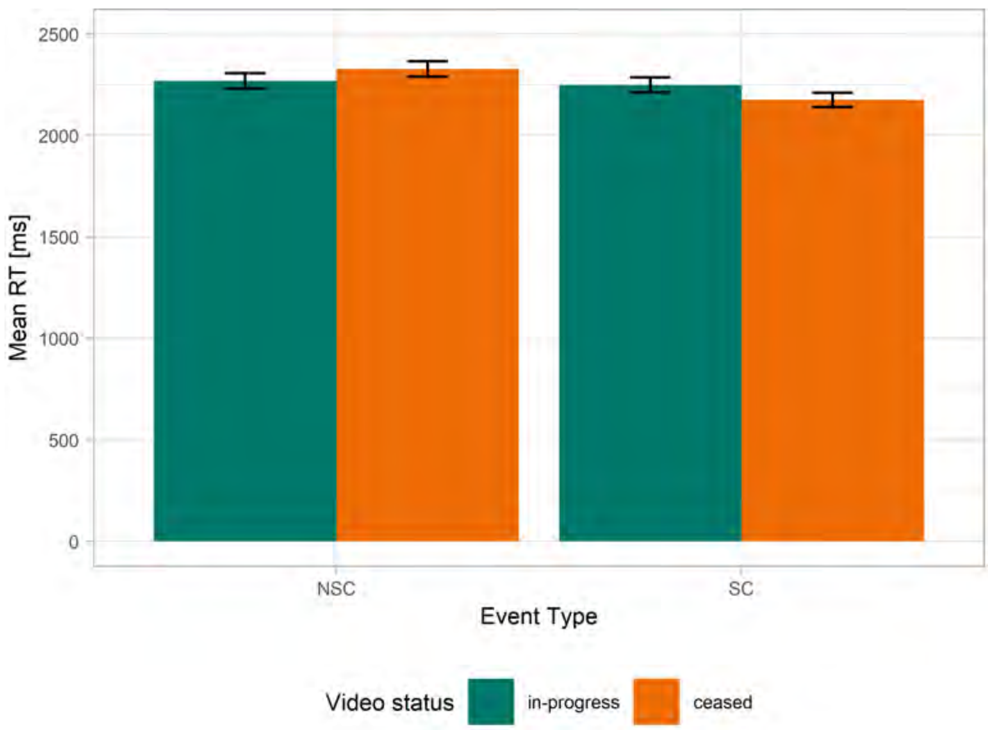


Figure 6: *Mean RTs (in milliseconds) for Event Type (NSC vs SC) and Video Status (in-progress vs ceased) ± 1 SEM.*

Table 6. RT analysis: Output of the model on log RTs of the correct responses to the video stills (lmer (logRT ~ Event Type * Video Status * Aspect + Corsi score + (Video Status | Participant) + (1 | Verb))

Fixed Effects	*β* Estimate	SE	*t*-Value
Intercept	7.595	0.15	50.600***
Event Type	0.026	0.01	2.543*
Video Status	-0.005	0.009	-0.592
Aspect	0.008	0.008	1.053
Corsi score	0.006	0.019	0.298
Event Type : Video Status	-0.028	0.008	-3.620***
Event Type : Aspect	-0.002	0.008	-0.272
Video Status : Aspect	0.004	0.008	0.547
Event Type : Video Status : Aspect	-0.0005	0.008	-0.075

***$p < 0.001$; **$p < 0.001$; *$p < 0.05$

Signal detection. No main effects were observed for sensitivity (d'). There was a significant interaction of Video Status and Event Type ($p < .001$, Fig. 7 left); SC events received more hits than false alarms in the ceased version ($M = 0.621$, $SD = 0.689$) compared to the in-progress version ($M = -0.255$, SD = 0.747). In contrast, NSC events received more hits than false alarms in the in-progress version ($M = 0.488$, $SD = 0.696$) as compared to the ceased version ($M = -0.275$, $SD = 0.697$).

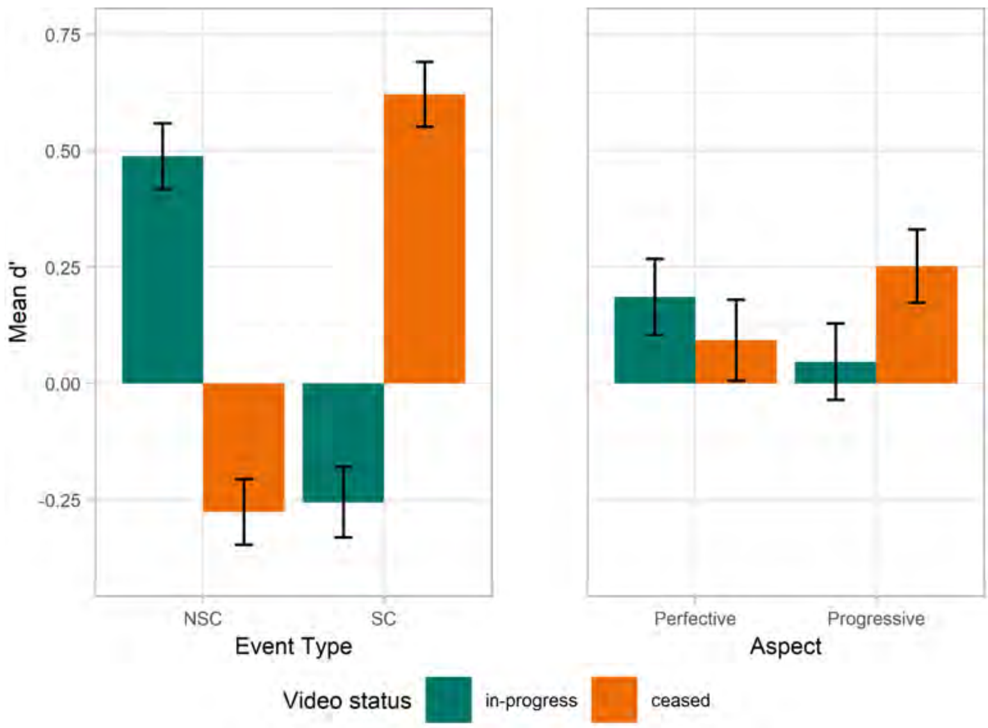


Figure 7: *Left: Mean d' scores for Event Type (SC vs NSC) and Video Status (in-progress vs ceased) ± 1 SEM. Right: Mean d' scores for Aspect (perfective vs progressive) and Video Status (in-progress vs ceased) ± 1 SEM.*

In addition, we observed an interaction of Video Status and Aspect (*p* = .037, Fig. 7 right). Specifically, d' scores were higher for ceased videos and sentences in the progressive (*M* = 0.252, *SD* = 0.781) compared to sentences in the perfective (*M* = 0.094, *SD* = 0.862). The d' scores for in-progress videos were higher when participants read a sentence in the perfective (*M* = 0.186, *SD* = 0.808) compared to the progressive (*M* = 0.047, *SD* = 0.812).

Table 7. d-prime analysis: Output of the model on sensitivity (lm (d-prime ~ Event Type * Video Status * Aspect + Corsi score)

Fixed Effects	*β* Estimate	SE	*t*-Value
Intercept	-0.074	0.147	-0.503
Event Type	-0.038	0.036	-1.075
Video Status	-0.028	0.036	-0.79
Aspect	-0.005	0.036	-0.139
Corsi score	0.029	0.019	1.535
Event Type : Video Status	0.41	0.036	11.504***
Event Type : Aspect	-0.018	0.036	-0.517
Video Status : Aspect	0.074	0.036	2.089*
Event Type : Video Status : Aspect	-0.002	0.036	-0.057

***$p < 0.001$; **$p < 0.001$; *$p < 0.05$

Further, for C, there was a main effect of Event Type (*p* = .012), suggesting participants were less conservative in their responses to SC events (*M* = -0.146, *SD* = 0.405) compared to NSC events (*M* = -0.044, *SD* = 0.398). There were no interaction effects.

Table 8. Criterion analysis: Output of the model on bias (lm (c ~ Event Type * Video Status * Aspect + Corsi score)

Fixed Effects	*β* Estimate	SE	*t*-Value
Intercept	-0.097	0.084	-1.16
Event Type	0.051	0.02	2.516*
Video Status	-0.016	0.02	-0.774
Aspect	0.002	0.02	0.074
Corsi score	0.0003	0.011	0.029
Event Type : Video Status	0.009	0.02	0.425
Event Type : Aspect	0.02	0.02	0.995
Video Status : Aspect	-0.005	0.02	-0.247
Event Type : Video Status : Aspect	0.018	0.02	0.862

***$p < 0.001$; **$p < 0.001$; *$p < 0.05$

5. Discussion

We set out to study the role of descriptive language, and specifically grammatical aspect as a marker for time in language, on event memory in two experiments. First, we studied the role of language for perception, serving as guiding input prior to perceiving an event (Experiment 1). Second, we studied language for encoding, where language was provided as a tool after experiencing an event (Experiment 2). We were interested in participants' memory for event endings of video clips of action events. In both experiments, we manipulated the event type (SC vs NSC), video status (in-progress vs ceased) and grammatical aspect (perfective vs progressive), and tested memory for video clip endings in a surprise recognition memory task.

Recall that for their accuracy measure, Santín et al. (2021) observed a main effect for event type in the verbal condition of their study, where participants were asked to describe events in video clips after watching them. Specifically, memory was better for SC events compared to NSC events. In Experiment 1, we also observed better memory accuracy for SC events compared to NSC events, though this effect was only approaching significance. The linguistic input provided before the videos may have led participants to construct an event model, which, based on verb semantics, activated state change (cf., Chapter 3, Experiment 2). This state change representation was then confirmed by the video. By contrast, in our Experiment 2, where descriptive language was provided after the videos, the main effect of Event Type, was absent, at least for the accuracy measure. However, we do see a main effect in the response bias measure C, in both experiments. In line with a general bias for goals in events (Lakusta & Carey, 2015; Lakusta & Landau, 2005), participants seem to have been less conservative in their responses to SC events. This suggests that SC events led participants to be more liberal in their decision-making. This could be due to the object state change depicted in the videos being overall highly salient, such that regardless of aspect and video status, this saliency 'anchored' SC events in people's minds. In other words, the degree of state change may be less relevant than the state change itself, though the effects for Event Type observed here remain subtle.

In line with our hypotheses, and in both experiments, we observed an interaction between Video Status and Event Type. Participants had more accurate memories for in-progress NSC events (as compared to in-progress SC events), and in responding to ceased SC events (as compared to ceased NSC events), which also showed in faster RTs and higher sensitivity scores. Note that the actions in the NSC events do not result in a state change in the object, meaning the object (e.g., a keyboard being typed on) looks the same throughout. In fact, the entire scenes do not substantially change visually throughout the event. By the end of the event (shown only in the ceased versions), the hands of the actor are withdrawn to signal arrival at the event boundary. In SC events, however, there is change across multiple dimensions. Here, too, the actor's hands are withdrawn at the end of the event, but the object has visually irreversibly changed. SC events might thus be better represented by the ceased videos, since both SC events and the ceased video status highlight the right event boundary characterised by the changed state of the object. By contrast, NSC events might be better represented by in-progress videos, since here both event type and video status lack a clear end state but match in their focus on ongoingness, and less on a resultant outcome or goal depicted through the object.

Crucially, we were interested in the role of aspect, which has previously been shown to affect event conceptualisation. Specifically, perfective aspect, highlighting event time corresponding to the right boundary, has been shown to activate visual representations of an object matching the end state of an event (Misersky et al., 2021; Madden-Lombardi et al., 2017). By contrast, progressive aspect defocuses event boundaries instead highlighting intermediate event stages (Madden & Zwaan, 2003; Madden & Ferretti, 2009). In Experiment 1, no effect of Aspect was observed, suggesting that the effect of aspect may be too subtle to guide attention for event memory in addition to event type. In Experiment 2, however, we did observe an interaction for the accuracy and d' measures, where Aspect interacted with Video Status: in-progress videos were remembered better with sentences in the perfective as compared to the progressive, whereas ceased videos were remembered better with sentences in the progressive as compared to the perfective. We hypothesised that the effect of aspect on memory could go into two directions: A match advantage, where, for example, the perfective cues for the resultant state, which then boosts memory for ceased video stills. Alternatively, there could be a mismatch advantage, such that when aspect and video status cue for different event stages, this would boost memory. A mismatch effect is indeed what we see in the results of Experiment 2. The difference in results for Aspect in the two experiments may be due to when language was introduced as a tool for memory.

The results in Experiment 1 would suggest that language (here: aspect) did not guide participants' attention to specific event features prior to watching it, with aspect showing no effect. In Experiment 1, language may have initially supported construction of an event model, which is then further specified by viewing the video, but showing no further effect on memory for perception. However, aspect seems to have affected memory accuracy after observing the event, for encoding. In Experiment 2, participants had already conceptualised the event after watching the video, and may have committed a representation of it to memory when they received the linguistic description. In that sense, the video could be seen as the major source of event-relevant information, and participants may have noticed the mismatch between what they have seen and the event status described by the aspect. This is in line with studies suggesting that language for encoding (i.e., after the event) can affect memory (Skordos et al., 2020). Note, that we only introduced language immediately before or after the events, similarly to Skordos et al. (2020), but not during probing where linguistic input has been shown to lead to misremembering (Loftus & Palmer, 1974). Instead of leading to misremembering, mismatching linguistic input in the present study actually improved memory performance, and future studies could further study the effect aspect may have an effect on memory retrieval.

In sum, we have observed that memory for events, and specifically event endings, is strongly influenced by event type. In our experiment, the event types differed with regards to whether an object was undergoing an irreversible change in state or not, which was indirectly represented by the verb semantics in the sentences. Our results are in line with previous findings by Santín et al. (2021) which also highlight the importance of change in the conceptualisation and memory of events. We have shown the accurate tracking of object states along the whole trajectory of change (cf. Intersecting Object Histories; IOH, Altmann & Ekves, 2019) influences how event endings are remembered. Tracking thus supports memory for event boundaries (Event Segmentation Theory; EST, Zacks & Swallow, 2007), with specific focus on the event boundary signalling the end of the event.

Acknowledgments

We thank Thijs Rinsma and Maarten van den Heuvel for their help with the implementation of the pre-tests and support setting up the studies, Miguel Santín for sharing his thoughts and his video stimuli with us, and Ksenija Slivac for advice regarding the analyses of the data.

6. Appendix

List of Descriptive Sentence Stimuli

	Condition	Subject	Perfective	Progressive	Object
1	NSC	She	bounced	was bouncing	a ball.
2	NSC	She	clapped	was clapping	hands.
3	NSC	She	cuddled	was cuddling	a teddy bear.
4	NSC	She	massaged	was massaging	a shoulder
5	NSC	She	measured	was measuring	a crate.
6	NSC	She	petted	was petting	a toy.
7	NSC	She	played	was playing	a xylophone.
8	NSC	She	poked	was poking	a duck.
9	NSC	She	polished	was polishing	a thermos.
10	NSC	She	pressed	was pressing	a lid.
11	NSC	She	read	was reading	a magazine.
12	NSC	She	rocked	was rocking	a doll.
13	NSC	She	salted	was salting	food.
14	NSC	She	showed	was showing	a picture.
15	NSC	She	smelled	was smelling	coffee.
16	NSC	She	spun	was spinning	a spinner.
17	NSC	She	sprayed	was spraying	water.
18	NSC	She	stirred	was stirring	soup.
19	NSC	She	studied	was studying	a map.
20	NSC	She	tested	was testing	a microphone.
21	NSC	She	twisted	was twisting	hair.
22	NSC	She	typed	was typing	on a keyboard.
23	NSC	She	watered	was watering	a plant.
24	NSC	She	waved	was waving	a napkin.
1	SC	She	bent	was bending	a paperclip.
2	SC	She	broke	was breaking	chocolate.
3	SC	She	connected	was connecting	dots.
4	SC	She	cracked	was cracking	a peanut.
5	SC	She	crumbled	was crumbling	a cookie.

6	SC	She	crumpled	was crumpling	a receipt.
7	SC	She	colour	was colouring	a square.
8	SC	She	drank	was drinking	juice.
9	SC	She	ate	was eating	a carrot.
10	SC	She	erased	was erasing	a circle.
11	SC	She	grated	was grating	cheese.
12	SC	She	inflated	was inflating	a balloon.
13	SC	She	mixed	was mixing	colours.
14	SC	She	painted	was painting	cardboard.
15	SC	She	peeled	was peeling	a banana.
16	SC	She	poured	was pouring	liquid.
17	SC	She	removed	was removing	nail polish.
18	SC	She	sliced	was slicing	an apple.
19	SC	She	mashed	was mashing	potatoes.
20	SC	She	swept	was sweeping	a table.
21	SC	She	tied	was tying	a knot.
22	SC	She	wiped	was wiping	sand.
23	SC	She	wrapped	was wrapping	a box.
24	SC	She	solved	was solving	a puzzle.

CHAPTER 5

THE POTENTIAL OF IMMERSIVE VIRTUAL REALITY FOR THE STUDY OF CROSS-LINGUISTIC DIFFERENCES IN OUR PERCEPTION OF THE WORLD[10]

Abstract

In everyday life, we actively engage in different events from a first-person perspective. However, experimental research on event cognition is often limited to relatively passive, third-person computer-based paradigms. In the present study we tested the feasibility of using immersive virtual reality (VR) in combination with eye-tracking technology on participants in active motion. Specifically, we used this methodology to study cross-linguistic differences in event cognition. Behavioural research has shown that speakers of aspectual and non-aspectual languages attend to goals (endpoints) in motion events differently, with speakers of non-aspectual languages showing relatively more attention to goals (endpoint bias). In the current study, native speakers of German (non-aspectual) and English (aspectual) walked on a treadmill across 3-D terrains in VR, while their eye gaze was tracked. Participants encountered landmark objects on the side of the road, and endpoint objects at the end of it. Based on previous research, we expected German speakers' fixation patterns to reflect an endpoint bias. Using growth curve analysis to analyse fixation patters over time, we found no differences between German and English speakers. The absence of cross-linguistic differences was also observed in behavioural control tasks with the same participants. The results are discussed in terms of second language experience: the exposure to and use of non-native aspectual languages (English and Dutch) in their everyday lives may have mitigated any endpoint bias in the German speakers. Overall, our VR setup is a promising method to study the effects of language on cognition in an ecologically valid setting.

[10]An adapted version of this chapter will be published as: Misersky, J., Peeters, D., & Flecken, M. (under review). The potential of immersive virtual reality for the study of event perception. *Frontiers in Virtual Reality.*

1. Introduction

In our day-to-day lives we are usually not confined to passive observation, but instead engage actively with the world around us. Making sense of the many events we experience day in, day out, is central to human cognition. An event can be characterised 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). As such, events are perceptual units with a specific spatial and temporal structure, which define the event's boundaries (i.e., its beginning and end). We segment the ongoing activity around us into discrete events by building and using structured memory representations or so-called event models (e.g., Radvansky & Zacks, 2014). When experiencing or viewing an event, we build a mental model of the encountered situation on the basis of such event models. Event models help us to make sense of the event we are currently experiencing. Properties relevant to an event model are the spatio-temporal framework (*where and when?*), the entities involved (*who and what?*), and the specific relations between, and properties of, these entities. Crucially, our knowledge of events and their properties is not solely based on physical involvement with them. It may also be shaped by our linguistic knowledge, as the way we process bottom-up information about events could be influenced top-down by language, and in particular by the language-specific structures provided by the languages we master and use (e.g., Gerwien & von Stutterheim, 2018).

The temporal contours of an event may for instance be made explicit linguistically through grammatical aspect marking. Specifically, progressive aspect (e.g. in English, "*She was dancing.*") can be used to highlight the inner temporal structure and continuousness (or progressiveness) of an event, such that the time discussed falls within the temporal boundaries of the event (Klein, 1994). In contrast, perfective aspect (e.g. in English, "*She danced.*") focuses on the event in its entirety, without specific details about the inner structure of the event itself. These different linguistic means of marking time in an event description may also affect how we perceive the different components that comprise an event (e.g., relevant objects and agents), and how we perceive the event's duration. Behavioural research on grammatical aspect has indeed found that events are perceived as having a longer duration when presented in the progressive form compared to the unmarked 'simple' form (Flecken & Gerwien, 2013, in Dutch). In English, progressive marking 'stretches' the time span of events, thereby honing in on the intermediate phases of an (ongoing) event (Hong et al., 2019; Madden-Lombardi et al., 2017; Matlock, 2011). Similar findings have been obtained for motion events (e.g., Anderson et al., 2013), which are also the focus of the current study. Generally, agents in events are thought to have desires and goals, which are brought about by actions such as motion (Magliano & Radvansky, 2001).

Importantly, languages differ in the ways in which aspect is expressed, and hence speakers of different language backgrounds may attend to different features within a scene depending on their language background. This hypothesis has been studied cross-linguistically, for instance by comparing speakers of aspectual languages (e.g., English, Modern Standard Arabic) to speakers of non-aspectual languages (e.g., German, Swedish) (von Stutterheim et al. 2012; Athanasopoulos & Bylund, 2013). Aspectual languages generally have a grammatical (and obligatory) means of expressing aspectual distinctions (such as ongoingness) through, for example, morphology. By contrast, in non-aspectual languages ongoingness can be expressed only in lexicalised ways, for example through adverbials or periphrastic constructions.

Interestingly, it has been observed that this difference in aspectual viewpoint expression also leads to differential gaze patterns and event descriptions across speakers of aspectual and non-aspectual languages. Watching video clips of motion events, participants have been asked to describe the events in one sentence, while their eye movements were tracked throughout (von Stutterheim et al., 2012). Cross-linguistic differences were found for video clips in which agents approached but did not reach a potential goal (e.g., a person running towards the train station, but the video ending before the person actually reached it). Specifically, the data revealed that speakers of non-aspectual languages (e.g., German) were more likely to mention a goal (i.e., the train station) in their descriptions, and also looked more often and longer towards the potential goal, compared to speakers of aspectual languages (e.g., English) (von Stutterheim et al., 2012). Interestingly, the results showed a pattern along a continuum: Within non-aspectual languages, Dutch speakers, for example, were less attentive to goals compared to German speakers, arguably because in Dutch ongoingness is more frequently expressed through lexical means compared to German. This highlights that any 'endpoint bias' may not be based solely on the grammatical categories (i.e., aspectual vs non-aspectual), but also on language use, in particular the frequency of use of the available aspectual devices in a given language.

In a follow up study using the same motion video clips and eye-tracking procedure, the two extremes observed previously—German as a non-aspectual language and Modern Standard Arabic (MSA) as an aspectual language—were compared (Flecken et al., 2014). The results showed a similar pattern: MSA speaking participants attended more to the ongoing motion (and agents) in an event compared to German speakers, who focused more on the potential endpoint of the motion they were seeing. Again, this difference was also reflected in their verbal description of the video clips: Speakers of a non-aspectual language expressed a potential goal more often than speakers of an aspectual language. Comparable results have been reported for speakers of German and English using causative events (Flecken et al., 2015).

This cross-linguistic difference in verbal motion event descriptions has further been replicated in a comparison of English (aspectual) versus Swedish (non-aspectual) speakers (Athanasopoulos & Bylund, 2013). Indeed, in the description of motion events, native speakers of Swedish verbally mentioned potential goals more often than native speakers of English. This finding was also confirmed by a similarity judgment task performed by a new set of participants. In this task, participants watched triads of video clips (adapted from von Stutterheim et al., 2012): First they saw clip A, then clip B, and finally clip X. Clip X was adjusted to always show intermediate goal orientation, whereas the alternates (A and B) showed low and high goal orientation (or vice versa) respectively. The task was to decide whether X was more similar to A or B. Swedish speakers opted for the alternative with high goal orientation more frequently than English speakers.

In sum, the above results suggest the following: As shown by their eye movement patterns and verbal event descriptions, speakers of non-aspectual languages (e.g., German, Swedish) are more likely to show a bias towards potential goals or endpoints when watching motion events. In contrast, speakers of aspectual languages (e.g., English) do not show such a bias, but are more inclined to attend towards features of a scene that signal ongoingness. This cross-linguistic difference is particularly apparent in verbal tasks, and seems to be less pronounced in non-verbal tasks (Athanasopoulos & Bylund, 2013). It is an open question to what extent previous findings generalise to everyday first-person perception of the world. To what extent do we actually perceive the world differently in everyday life as a function of our language background? Following up on previous work, we aim to extend the study of endpoint bias in motion events by comparing German and English native speakers in a dynamic and immersive 3-D setting.

Over the past decades, experimental research into human language and cognition has greatly benefited from the use of computer monitors to display carefully developed experimental stimuli to participants. The use of strict experimental designs implemented through computer paradigms has indeed allowed for the advancement of cognitive and psycholinguistic theory in unprecedented ways. Recently, however, in several subfields of the experimental study of language, perception, and cognition, more and more attention is devoted towards combining high experimental control with high ecological validity (e.g., Blanco-Elorrieta & Pylkkänen, 2018; Hari et al., 2015; Knoeferle, 2015; Peeters, 2019; Willems, 2015). Crucially, unlike in the typical traditional experimental setup that places individual participants in front of a computer monitor, in real-life settings 'people are participating in the events of their world, and they do not only serve as passive observer' (Hari et al. 2015, p. 184). In this view, traditional studies and their interpretations are limited by employing passive and highly controlled experimental stimuli. Specifically, it is unclear whether the theories and models that were built on the basis of traditional, strictly controlled computer paradigms generalise to people's everyday behaviour in the real world.

As such, interactive settings, which combine high ecological validity with high experimental control, could provide us with a clearer understanding of human cognition as active under natural circumstances (Pan & Hamilton, 2018; Peeters, 2019). It is an open question, though, whether novel methods will yield findings that differ from the findings obtained in traditional paradigms. For instance, recent work suggests that using a third-person or first-person perspective in event video clips may not necessarily have a strong influence on how those events are perceived (Magliano et al., 2014). Nevertheless, further study is required to test whether an individual's language background and, more specifically, the grammatical affordances of their native language, influence how they perceive events in the real world from a first-person perspective in general.

1.1. The Present Study

In the current study, we therefore follow up on previous work by utilising 3-D projections in a cave automatic virtual environment (CAVE) in which participants can actively engage in motion themselves (Cruz-Neira et al., 1993). This creates a realistic setting to reflect day-to-day situations, allowing for the direct study of potential cross-linguistic differences in first-person perceptual experiences while navigating through space. Using the CAVE setup, we are able to implement Hari et al.'s (2015) advice to get participants to engage with more naturalistic stimuli in motion. This allows us to step away from computer-based tasks while still exerting the required control over our stimuli and collecting reliable data.

Recently, a number of different studies have successfully used immersive VR as a reliable method to study fundamental aspects of human language and cognition (for an overview, see Peeters, 2019). For instance, VR and eye tracking have been combined to gain clearer insights into how listeners may predict upcoming words for visual objects in visually rich settings (Heyselaar et al., 2018). In addition, VR has been successfully used to study language-related topics such as syntactic priming, audiovisual integration, bilingualism, and the role of eye gaze in social interaction (e.g., Cañigueral & Hamilton, 2019; Heyselaar et al., 2015; Peeters & Dijkstra, 2018; Tromp et al., 2018). In these studies, the virtual environment typically had a substantial degree of dynamicity, but participants themselves remained relatively static throughout the experiment. In a next step, we here aim to move away from largely static 3-D settings, and put participants into motion in a moving VR environment. Specifically, we will place native speakers of German (a non-aspectual language) and English (an aspectual language) on a treadmill in an immersive 3-D environment to test whether their first-person perception of the world differs as a function of their language background.

Our aims with this more ecologically valid experimental setup were twofold: First, we set out to test the feasibility of combining immersive VR with eye tracking in moving participants. If reliable data can be collected in such a setup, this will open up a wide range of possibilities for future studies investigating the relation between language and perception, for instance in the domain of (motion) event cognition. Second, the current setup allowed us to collect more fine-grained temporal information—that is, how gaze patterns will develop over the time course of an ongoing event in which the participant is actively involved and experiences their environment from a first-person perspective. This will add to the existing literature in the fields of psycholinguistics and event cognition in which eye tracking was used (e.g., von Stutterheim et al., 2012; Flecken et al., 2014; 2015). In line with previous research, we employed a design in which participants encountered landmarks as well as potential endpoints (see below), and were interested in their looking behaviour to the latter as a function of their native language background. In particular, we expected Germans to fixate on endpoints more compared to English speakers. To keep the setup as similar to real-life motion as possible, participants were not prompted to verbalise what they were experiencing during the VR task. To be able to compare this novel setup to previous work, we also utilised verbal control tasks on event cognition, which have resulted in robust cross-linguistic differences in previous work (see below).

2. Methods

Participants

Twenty-four speakers of German (15 female, 20-31 years of age, *M* = 22.92, *SD* = 2.84), and 24 native speakers of English (14 female, 18-35 years of age, *M* = 23.92, *SD* = 4.6) participated in this experiment.[11] Participants were invited via the online participant database of the Max Planck Institute for Psycholinguistics, and through advertising on flyers and social media. They gave written informed consent prior to their participation and received a monetary compensation for their time (10 Euros per hour). The experiment was approved by the ethics committee of the Social Sciences Faculty of Radboud University, Nijmegen, The Netherlands.

After completing the experiments, we asked participants to provide information on their educational and language background. Specifically, we asked about other countries they have lived in (for a minimum of three months) and which other language, apart from their native language, they were most proficient in. In addition, we asked them to rate their ability of speaking, writing, comprehension and reading in their most proficient L2 (poor, sufficient, good, very good, or excellent). Overall, English native speakers came from a greater variety of home countries (incl. Australia, Canada, Luxemburg, Indonesia, UK, US, and Trinidad) compared to German speakers (Germany). English speakers frequently reported poor or sufficient skills in their most proficient L2 (e.g., French, Spanish, German) whereas German speakers most frequently named English and Dutch as their most proficient L2 with skills predominantly rated as very good and excellent. Note also that the majority of all participants were international students at Radboud University, where study programmes were held in English, meaning that

[11]In addition, one English speaker participated but was excluded, as they were raised in both English and German. Further, nine German-speaking participants took part but were excluded from further analysis due to technical issues resulting in the incorrect presentation of stimulus lists (for further exclusion in the behavioural tasks see the Results section).

the German speakers were often exposed to a language other than their L1, and frequently used this/these language(s) in daily life, too.

VR Material Selection and Trial Setup

In the VR part of the experiment, participants walked through a total of 48 trials on a treadmill. Four 3-D road types were designed: a parkland lane, an urban road, a sandy countryside path, and a forest trail (see Fig. 1). Fifty-two unique 3-D objects were selected for a pre-test. Using a paper-and-pencil questionnaire, 18 native German speakers (9 female, 25-41 years of age, *M* = 30.11, *SD* = 4.68) and 20 English speakers (10 female, 21-37 years of age, *M* = 27.8, *SD* = 3.64) first provided a name for a 2-D picture of an object, and then rated its prototypicality on a 7-point scale (1 = not prototypical at all, 7 = very prototypical). None of the participants took part in the main experiment. For the experimental trials of the main experiment, the 48 objects with highest name agreements and prototypicality ratings within each language were chosen. Forty of the objects were specifically designed for this experiment by a graphics designer using Autodesk Maya software. The remaining objects (*N* = 8) were taken from the standardised database of 3-D objects provided by Peeters (2018).

Twenty-four of the trials were experimental trials, in which participants always encountered two 3-D objects, one on the side of the road (landmark, LM) and one at the end of the road (endpoint, EP). Unbeknownst to participants, each trial was split into three phases. In Phase 1, only the LM was visible. As the trial continued, participants entered Phase 2, in which both LM and EP were in view. As they passed the LM, participants entered Phase 3, in which only the EP was visible. Half of the trials stopped before the EP was reached (short trial), whereas in the other half, the trial stopped as they arrived at the EP (long trial). This ensured that participants could not anticipate whether or not they would actually reach the goal (EP object), similar to previous behavioural studies based on video clips (von Stutterheim et al, 2012; Flecken et al., 2014). Importantly, previous cross-linguistic differences were obtained only for events in which actors did not reach a goal; this was the critical condition leaving room for variability in specifying and looking at potential goals. Here, Phase 2 would correspond to a situation in which a potential goal is visible, but whether or not it would be reached was unclear yet, regardless of whether trials were long or short.

The objects were matched to appear always with the same road type (see Appendix for a list of all objects and the corresponding road types). For each road type, the items were balanced across participants such that an object's presentation was balanced across long and short trials, in their position in LM or EP location, and whether they appeared on the left or right side of the road in the case of LM. This resulted in eight lists across participants, in which trials were pseudo-randomised for each participant. In the filler trials (*N* = 24) participants encountered a virtual agent crossing the road in front of them. Four virtual agents (two female) were adapted from existing stock avatars produced by WorldViz (Vizard, Floating Client 5.4, WorldViz LLC, Santa Barbara, CA). See Fig. 1 for an example of the trial setup and stimuli.

Figure 1: *Left to right: Examples of the four road types and objects within them. Top left: A parkland lane with LM (a tractor) on the right; top right: an urban road with LM (bus stop) on the right. Bottom left: A sandy countryside path with EP (camper trailer) at the end of the road; bottom right: a forest trial with a virtual agent crossing the road.*

Apparatus

The CAVE System. The CAVE system is made up of three screens (255 × 330 cm, VISCON GmbH, NeukirchenVluyn, Germany), which were arranged to the left, to the right and in front of the participant, as illustrated in the in Fig. 2. Each of the screens was illuminated through a mirror by two projectors (F50, Barco N.V., Kortrijk, Belgium). The two projectors showed two vertically displaced images, which overlapped in the middle of each screen, meaning the display was only visible as the combined overlay of the two projections. Infrared motion capture cameras (Bonita 10, Vicon Motion Systems Ltd, UK) and the Tracker 3 software (Vicon Motion Systems Ltd, UK) allowed for optical tracking. The infrared cameras detected the positions of retroreflective markers mounted onto 3-D glasses by optical–passive motion capture (see below for further details). A total of ten infrared cameras were placed around the edges of the screens in the CAVE: Six cameras were positioned at the upper edges, and four cameras at the bottom edges. All cameras were oriented toward the middle of the CAVE system, where the participants were located during testing. The positions of a subset of the cameras are indicated in Fig. 2. The resolution of the CAVE system was 2560x1956 pixels per screen and the refresh rate was 60 Hz, which allowed for the glasses and CAVE system to be in sync. The CAVE setup is described in further detail in Eichert et al. (2018).

Unlike earlier studies using this CAVE lab, a treadmill was placed in the centre of the system, such that during walking, all three screens covered participants' entire horizontal visual field. The eyes of the participant were approximately 180 centimetres away from the middle screen.

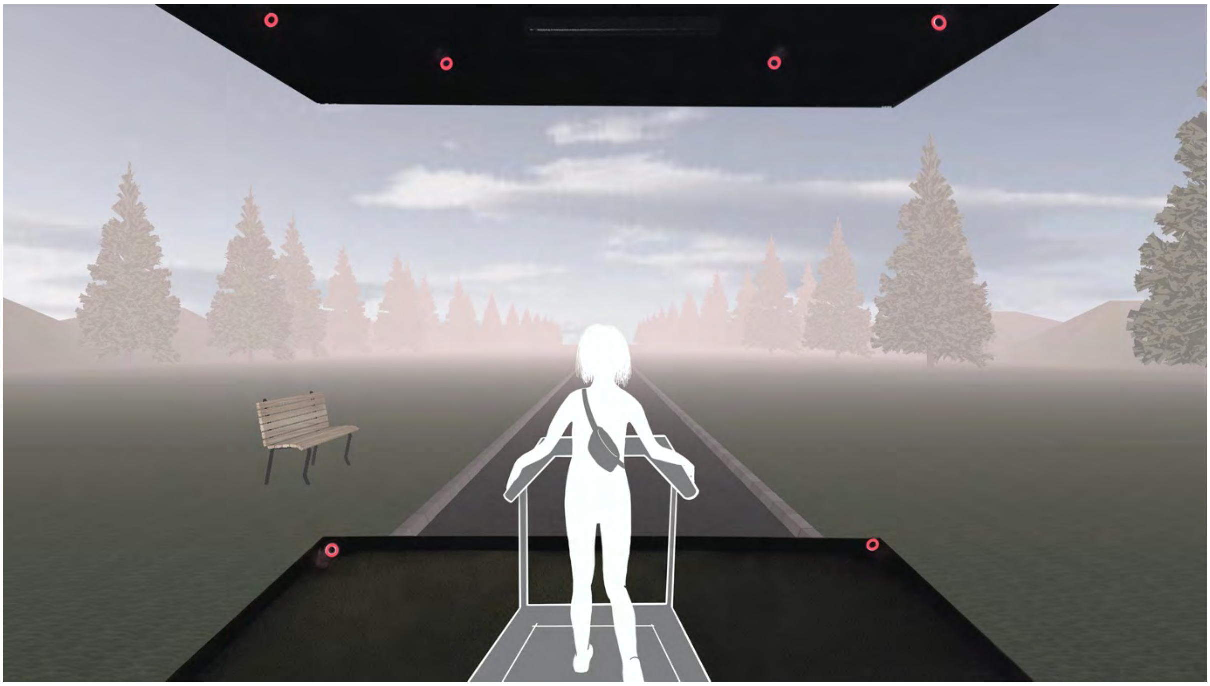


Figure 2: *Illustration of the CAVE setup with its three screens. Participants walked on a treadmill through a variety of dynamic virtual environments, encountering both landmark (here: a bench) and endpoint objects along the way while their eye movements were continuously tracked and recorded. A subset of the infrared cameras is marked red.*

From the control room, the experimenter could see the participant and the displays on the screens through a large window behind the participant (thus facing the central CAVE screen, similar to the view depicted in Fig. 2).

The experiment was programmed in Python, and run using 3-D application software (Vizard, Floating Client 5.4, WorldViz LLC, Santa Barbara, CA). To allow for a realistic experience, we made sure that moving forward a 'virtual' metre inside the VR environment corresponded to moving forward a 'real' metre on the treadmill. Note that object sizes were dynamic, i.e., perceived object size changed throughout a trial as the participant advanced on the virtual terrain – as in real life objects became relatively larger when participants approached them. To further enhance the naturalness of the walking experience, bird and wind sounds were presented through speakers located in the CAVE (Logitech Surround Sound speaker system Z 906 5.1).

Eye Tracking. Eye tracking was performed using glasses (SMI Eye-Tracking Glasses 2 Wireless, SensoMotoric Instruments GmbH, Teltow, Germany), which combine the recording of eye gaze with the 3-D presentation of VR through shutter glasses. The recording interface was based on a tablet (Samsung Galaxy Note 4), which was connected to the glasses by cable. The recorder communicates with the externally controlled tracking system via a wireless local area network (WIFI), allowing for live data streaming. The glasses were equipped with a camera for binocular 60 Hz recordings and automatic parallax compensation. The shutter device and the Samsung Galaxy 4 tablet were placed in a small shoulder bag that participants carried on their back during walking on the treadmill (see Fig. 2). Gaze tracking accuracy was estimated by the manufacturer to be 0.5° over all distances. The delay of the eye-tracking signal (i.e., the time it takes for the eye-tracking coordinates to reach the VR computer and its output file) was estimated to be 60 ± 20 milliseconds. In addition, we combined the eye tracking with optical head-tracking. Optical head-tracking was accomplished by placing spherical reflectors on the glasses. We were thus able to identify the exact location of the eye gaze in three spatial dimensions (X, Y and Z). This in turn allowed for an immersive experience in the VR, as the presentation of the 3-D world moved in accordance with the participants' (head) motion on the treadmill. In order to achieve smooth 3-D presentation, the 3-D eye-tracking glasses were equipped with reflectors (three linked spheres) magnetically attached to both sides of the glasses, which worked as passive markers that were detected by the infrared tracking system in the CAVE. The tracking system was trained to the specific geometric structure of the markers and detected the position of the glasses with an accuracy of 0.5 millimetres.

Calibration of the eye-tracker was carried out in two steps. In an initial step, general tracking of the pupils was tested in the control room using the Samsung Galaxy 4 tablet. If this was successful, shutter device and tablet were stowed into the shoulder bag, and a second calibration step was carried out in the 3-D environment as the participant was walking on the treadmill. For this step, we used a virtual test scenery, resembling a tea house in which three differently coloured spheres were displayed in front of the participants. The position of the tree spheres differed in all three spatial coordinates. Participants were asked to look at the three displayed spheres successively, which the experimenter communicated to the participants via the microphone. The computer software computed a single dimensionless error measure of the eye-tracker combining the deviation in all three coordinates for all three spheres. This second calibration step was repeated until a minimal error value (< 5° difference between the invisible vector between the shutter glasses and the centre of each sphere and the invisible vector between the shutter glasses and where a participant's fixation was actually estimated to be by the system), and thus maximal accuracy, was reached.

Regions of Interest. To determine target fixations, we defined individual 3-D regions of interest (ROIs) around each object in the virtual space. ROIs were defined by a rectangular prism that enclosed the object. The X (width) and Y (height) dimensions of the ROI were adopted from the frontal plane of the object's individual bounding box, facing the participant. In the experiment software, eye gaze towards an object was detected if the line of sight collided with an object's ROI in the scene. In other words, the eye-tracking software automatically detected when the eye gaze was directed at one of the ROIs and coded the information online in the data stream. Note that the dimensions of width, height and depth of a given ROI do not change, but the position of the participant in the virtual world changes because of the forward (Z-axis) motion. This can be thought of as an analogy to motion in the real world: The absolute values of height, width and depth of, for example, a parked car do not change; the car is always the same size. However, as you are approaching the parked car, your line of sight adjusts so you can see the top, front or side of the car.

Data Processing. Based on the inspection of the output data files, our data were cleaned before further processing. Each row in the output signified one frame of ~16.6 milliseconds. In some cases, the data showed timestamp duplicates. If a timestamp occurred more than once within a participant and trial, all frames with the duplicate were marked. Further, freezing during eye tracking also incidentally occurred, meaning that the coordinates for the X, Y and Z planes showed duplicates as well. When simultaneous freezing across all three coordinates was observed more than once within a participant and trial, all frames with the duplicate were marked. Marked frames were considered unreliable data points. Overall, 14.96% of the data were affected by duplicates in Phase 1, 11.22% in Phase 2, and 26.07% in Phase 3. The data were binned prior to analysis, with each bin containing three frames. Any bin containing at least one marked frame was excluded from the final analysis. A fixation was then defined as saccades towards an ROI for at least six frames or ~100 milliseconds. Shorter saccades were considered as unlikely to represent a fixation (cf. Eichert et al., 2018).

VR Eye-Tracking Data Analysis. The eye-tracking data obtained in the VR experiment was analysed using growth curve analysis (GCA) on the cleaned data (see above). GCA uses a linear mixed regression approach and has been successfully used for visual world paradigms (VWPs) before (see Mirman, 2014). GCA can give insights into whether looking behaviour differs between groups or items within a given time-window, and thus allowed us to see whether our language groups differed with regards to their fixations on EPs in the VR scenes. Thus, we focused our analysis on the phases when the EP is visible, namely Phases 2 and 3. Unlike conventional VWPs, our study used a free viewing setup. As such, there were no fixed time points from which we were able to restrict our time-windows of interest. Small changes –depending on the setup– can impact the results and their interpretation (Peelle & Van Engen, 2021). That said, there are no specific recommendations for the choice of time-window length, and transparency regarding the chosen approach is hence key. To keep the bias in choosing the time-windows to a minimum, we opted to use the same time-window length for both Phases 2 and 3. Combining visual inspection (i.e., when fixations start diverging in Phase 2) and information about Phase length in our setup (i.e., the maximal length of Phase 3 in short trials being five seconds), we opted to restrict our analysis to the last five seconds of Phase 2, and the first five seconds of Phase 3[12].

GCA uses polynomial orthogonal time terms as predictors in the linear mixed regression model, which describe the shape of the fixation curves in our time-windows of choice. For both Phase 2 and 3 we chose third-order polynomial time terms as predictors in the respective models. Each time term describes a different aspect of the fixation curve (cf. Sauppe, 2017): $Time^1$ (linear) describes the angle of the fixation curves. $Time^2$ (quadratic) describes the rate of increase or decrease. $Time^3$ describes earlier or later increases or decreases of the fixation curves. We ran generalised linear mixed effect models on the EP hits by Language (treatment-coded) interacting with the Time predictors as fixed effects. In addition, we introduced a fixed effect of a Nuisance predictor (Sassenhagen & Alday, 2016), which took into account all EP hits in the previous bin to use as a predictor for the current bin in order to reduce temporal autocorrelation in the continuous eye-tracking data. The maximal random effects structure allowing for convergence was used for all models (see full models in Tables 1 and 2). All models were computed using the lme4 package (Version1.1-21; Bates et al., 2015) in R (Version 3.6.0; R Core Team, 2019).

In line with our research question, interactions of any of the time terms with our Language predictor would indicate differences in EP fixation behaviour between the German and English speakers.

Procedure

After receiving written information about the experiment and giving consent, every experimental session started with the VR experiment. Participants put on the VR glasses, which were fastened with a drawstring strap. The calibration was then carried out, and when successful, all equipment powering the glasses was stored in a small shoulder bag, which the participant was

[12]The use of time-windows of five seconds was chosen for the following reasons: In line with our interest in views towards the EP, we wanted to analyse data for those time-windows in which participants had a free choice to look towards either the EP or something else. Recall that in Phase 2, participants had this choice throughout, whereas in Phase 3, participants had free choice of looking or not looking towards the EP only in the beginning of the Phase (coinciding with the end of short trials). We opted to take an objective and conservative approach to the data analysis, and thus wanted to use the same time-window length for both Phases 2 and 3. Please consult the Appendix (5.2.) for an additional analysis of Phase 2 using a shorter time-window and yielding comparable results.

instructed to wear throughout the experiment. The participant was then led into the room with the CAVE system and asked to stand on the treadmill. The second calibration step was then carried out, ensuring accurate tracking of the eye gaze and simultaneously checking whether the participant could see the projections in 3-D. The treadmill was turned on and set to a fixed, comfortable walking speed (~3km per hour) to match the presentation of the moving environment. The experimenter then explained the task of the experiment, which was to walk on the treadmill and listen out for bird sounds. Upon hearing a bird, the participant had to press a button on the handle of the treadmill. The participant experienced four practice trials (one of each road type, LM on the left half of the trials, equal amounts of long and short trials) to get used to the setup, and had the opportunity to ask questions afterwards. The experiment proper then started with another calibration in the CAVE. After half the trials, the participant had a self-paced break but had to remain on the treadmill. The experimenter re-checked the calibration, and re-calibrated where necessary, then started the second half of the experiment when the participant was ready. After completion of the VR experiment, the experimenter helped the participant to step off the treadmill and remove the VR glasses.

Three computer-based tasks followed, all of which were programmed using Presentation software (Neurobehavioral Systems).

First, participants carried out an object recognition memory task as a test of whether they paid attention during the VR experiment. On a white screen, participants saw a fixation cross, then an object for 1200 milliseconds, which was followed by the question whether they had seen it previously in the VR experiment. Participants had to indicate their decision via a button press on a button box. Overall, each participant saw 24 previously seen objects (12 as EPs, 12 as LMs), and 24 new objects. The order of objects was randomised. High accuracy overall in this task was hypothesised to indicate participants had paid attention to the objects in the task.

Second, participants performed an event description task (Flecken et al., 2014; von Stutterheim et al., 2012), in which on each trial they saw a fixation cross followed by a video clip. They were asked to describe what is happening in the clip using a single sentence. Participants were instructed to speak into a small microphone, and were allowed to speak as soon as they felt ready to describe what is happening. They were asked not to focus on details (e.g., colours, backgrounds, such as “the sky is blue”) and just on the event itself. A blank followed, and the fixation cross re-appeared, upon which participants were able to start the next clip. Overall, each participant saw 50 video clips in a randomised order. In line with previous work, our clips of interest were those in which a potential goal was not reached. For those clips (N = 10), German speakers were expected to mention EPs more often compared to English speakers.

Third, participants were instructed to carry out a similarity judgment task (adapted from Athanasopoulos & Bylund, 2013, their Experiment 2a). In each trial, they saw a triad of three consecutive short video clips, labelled A, B and X. Clip X always showed intermediate goal orientation, meaning there was a possible endpoint but no arrival was shown. The alternates (A and B) showed low (i.e., motion along a trajectory, no immediate endpoint) and high goal orientation (i.e., arrival at endpoint was shown) respectively. Their task was to judge whether clip X was more similar to clip A or to clip B. To indicate their decision, they had to press one of two buttons. In total, they saw 38 triads in a randomised order. German participants were expected to show higher endpoint bias (i.e., rating X as more similar to the high goal alternate) compared to English speakers.

Lastly, participants filled in a paper questionnaire regarding their language and educational background (see above for details). The experimenter thanked them for their time and debriefed them regarding the purpose of the study.

All interactions between the experimenter and the participants were carried out in the participants' native language. An experimental session took between 90 and 120 minutes.

3. Results

3. 1. VR Experiment

Phase 2: Both LM and EP Are Visible.

Fig. 3 shows participants' fixations to the EP in all of Phase 2. For the GCA analysis, we focused on the last five seconds of Phase 2, the results of which can be found in Table 1.

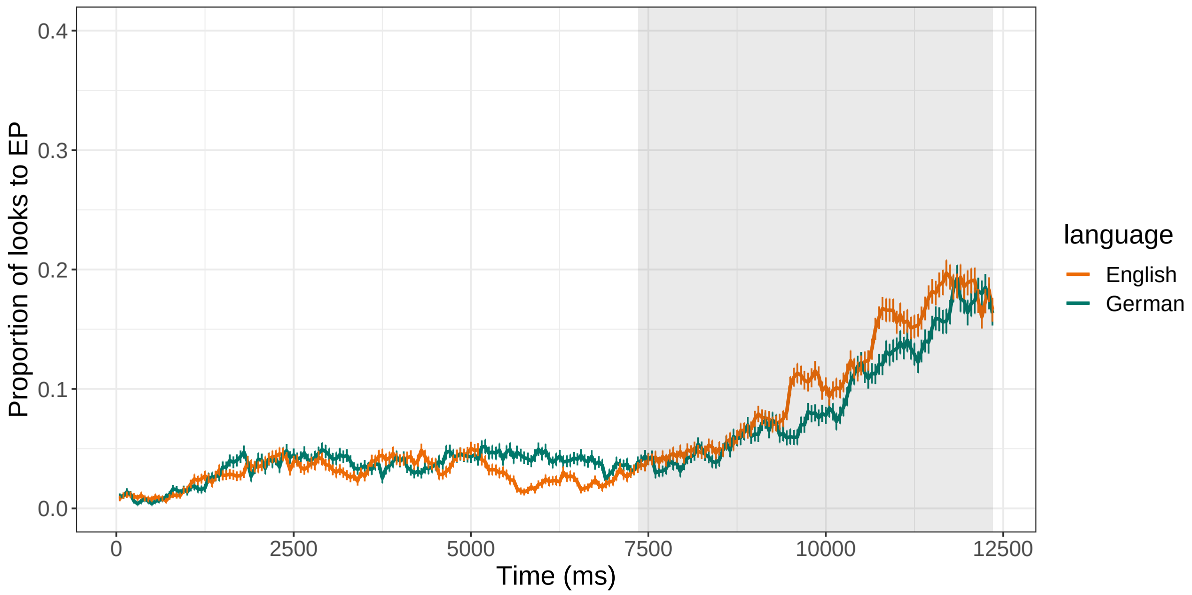


Figure 3: *Proportion of looks (fixations) to the EP in Phase 2 (both LM and EP visible) by Language, with the analysis time-window of five seconds highlighted in grey.*

The significant main effect for $Time^1$ ($p < 0.001$), here with a positive β Estimate, describes an overall increase in looks towards the EP over time, which is to be expected as participants approached the EP in this phase. The negative β Estimate of the main effect for $Time^3$ ($p < 0.001$) describes the "S-shape" of the fixation curve, which is not pronounced. No interactions between the orthogonal Time terms and Language were observed, suggesting no differences in looking behaviour (EP fixations) across the two Language groups in this Phase.

Table 1. Output of the model on EP fixations in Phase 2 (glmer (Hits ~ Language * ($Time^1$ + $Time^2$ + $Time^3$) + Nuisance + (1 | Item-pair) + (1 | Participant))

Fixed Effects	*β* Estimate	SE	*z*-value
Intercept	-4.057	0.185	-21.933***
Language Eng	-0.048	0.229	-0.21
$Time^1$	3.551	0.216	16.441***
$Time^2$	-0.048	0.215	-0.224
$Time^3$	-0.762	0.21	-3.624***
Nuisance	1.837	0.009	210.835***
Language Eng : $Time^1$	-0.049	0.294	-0.168
Language Eng : $Time^2$	-0.463	0.292	-1.586
Language Eng : $Time^3$	-0.051	0.287	-0.178

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

Phase 3: Only EP Is Visible

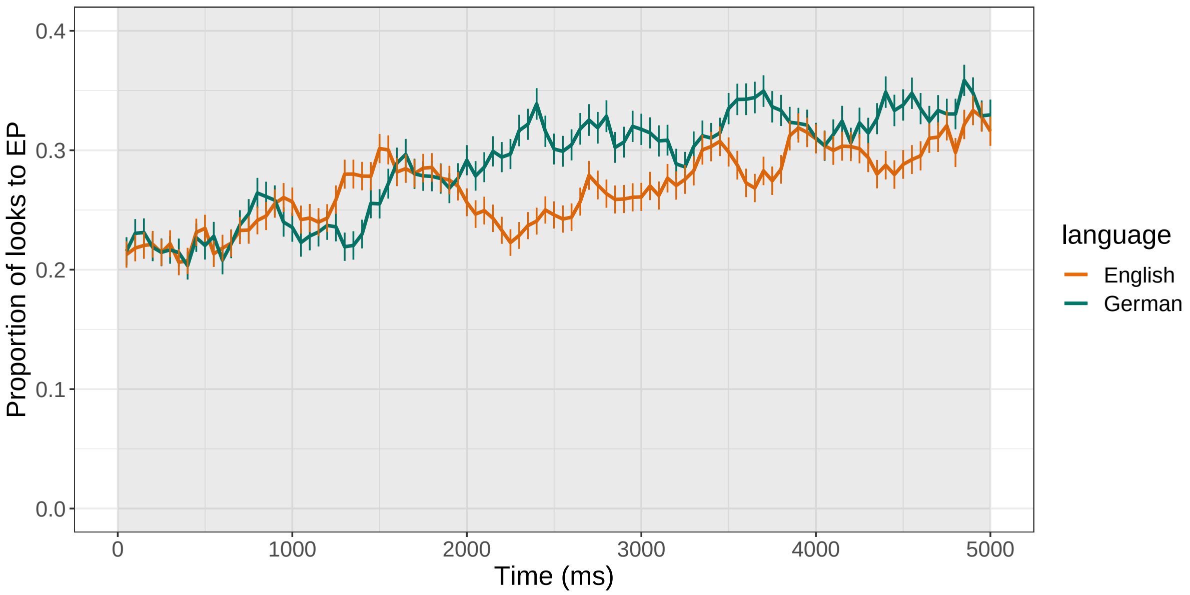


Figure 4: *Proportion of looks (fixations) to the EP in Phase 3 (only EP visible, collapsed for long and short trials) by Language, with the analysis time-window of five seconds highlighted in grey.*

Table 2 shows the results for the GCA analysis in the first five seconds of Phase 3. Like in Phase 2, we observed a significant main effect of $Time^1$ ($p < 0.001$) with a positive *β* Estimate, meaning overall looks towards the EP increased over time as to be expected. The negative *β* Estimate of marginally significant main effect of $Time^2$ ($p = 0.032$) reveals that the rate of increase was low. Most importantly, there was a marginal interaction between English and $Time^2$ ($p = 0.037$) with a positive *β* Estimate value, suggesting English speakers showed a slightly stronger rate of increase in fixations towards the Endpoint object in this Phase.

Table 2. Output of the model on EP fixations in Phase 3 (glmer (Hits ~ Language * ($Time^1$ + $Time^2$ + $Time^3$) + Nuisance + (1 + $Time^1$ + $Time^2$ | Item-pair) + (1 + $Time^1$ + $Time^2$ + $Time^3$ | Participant))

Fixed Effects	β Estimate	SE	*z*-value
Intercept	-1.826	0.141	-12.929***
Language Eng	-0.161	0.162	-0.996
$Time^1$	1.438	0.258	5.576***
$Time^2$	-0.543	0.253	-2.148*
$Time^3$	0.187	0.218	0.858
Nuisance	2.585	0.009	274.799***
Language Eng : $Time^1$	-0.294	0.32	-0.918
Language Eng : $Time^2$	0.678	0.325	2.084*
Language Eng : $Time^3$	-0.139	0.307	-0.453

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

3. 2. Object Memory Task

An analysis of the object memory task was performed for the 24 German speakers and 24 English speakers, who successfully took part in the main VR experiment. Both English (*M* = 0.860, *SD* = 0.347) as well as German-speaking (*M* = 0.859, *SD* = 0.349) participants' accuracy was well above chance in this task, with no significant difference (p = 0.9) between the two groups, suggesting all participants were attentive to the objects in the VR scenes.

3. 3. Event Description Task

Based on the data that was included in the VR analysis, technical errors or reported familiarity with the event description task, a number of data sets were excluded. The analysis of the event description task was thus based on 22 German speakers and 24 English speakers, and focuses on the subset of the trials in which an EP was not reached (*N* = 10). A score of ten indicated an EP was mentioned in all clips, whereas a score of zero indicated no EP was mentioned for any of the clips. Means of EP mentions were similar across English (*M* = 4.167, *SD* = 1.993) and German speakers (*M* = 4.091, *SD* = 1.925). We ran a logistic regression model on the binomial data of score (0 = EP not mentioned within a trial, 1 = EP mentioned within a trial), with a fixed effect for Language, and random effects for Participant and Video Clip. Results indicate there were no differences in EP mentions between the Language groups (Intercept: β Estimate = -0.365, SE = 0.744, *z*-value = -0.491, p = 0.624; Language: β Estimate = 0.056, SE = 0.390, *z*-value = 0.145, p = 0.885).

3. 4. Similarity Judgment Task

Based on the data which was included in the VR analysis, technical errors etc., a number of data sets were excluded. The analysis of the event description task is thus based on 21 German speakers and 23 English speakers. A score of one indicated that participants always judged all clips X to be more similar to the high-goal orientation alternate, whereas a score of zero meant that participants judged all clips X to be more similar to the low-goal orientation alternate. Means for a bias towards matching clips based on their degree of endpoint-orientation were similar across English (*M* = 0.272, *SD* = 0.445) and German speakers (*M* =

0.247, *SD* = 0.431). We ran a logistic regression model on the binomial data of Bias (0 = no bias, 1 = strong bias) by Language, with random effects for Participant and Video Clip. Results indicate there were no differences in similarity judgments between the Language groups (Intercept: β Estimate = -1.197, SE = 0.369, *z*-value = -3.240, *p* = 0.001; Language: β Estimate = 0.169, SE = 0.1417, *z*-value = -1.191, *p* = 0.234).

4. Discussion

The present study had two aims. First, following up on behavioural studies (e.g., von Stutterheim et al., 2012; Flecken et al., 2014), we tested for the generalisability of previous findings in the domain of cross-linguistic event perception to settings of first-person perception and self-propelled motion. Second, we set out to establish whether immersive VR can be used with moving participants in dynamic 3-D settings to study aspects of human language and perception.

To what extent do we perceive the world differently as a function of the grammatical affordances provided by our native language? As no cross-linguistic differences in looking behaviour were observed, our study did not replicate previous differences between native German and native English speakers in the proportion of looks to endpoints. While previous studies suggest German speakers look more towards endpoints than English speakers do, we were not able to observe this pattern in our eye-tracking data. It could be the case that participants overall engaged with the 3-D stimuli to an extent that could have washed out any effects. After all, this was the first time for all participants to have moved around in a virtual environment in this non-verbal part of the experiment. However, our two main control tasks both relied on language. The event description task explicitly asked participants to verbalise what they were observing. The event similarity judgment task required participants to hold the video events in mind, and participants may have used internal language to do so (cf. Athanasopoulos & Bylund, 2013). Interestingly, both control tasks also showed no cross-linguistic differences.

To what extent can virtual reality be used as a tool to investigate first-person perception of the world in rich and dynamic settings? With our second aim in mind, the lack of effects observed is unlikely due to VR methodology since the results from all parts of the present study show no differences on the basis of language background. In addition, our data suggest that a change from third to first person perspective is unlikely to affect our perception of event endpoints (cf. Magliano et al., 2014), as the VR study (first-person experience) as well as the control tasks (third-person experience) yielded similar results.

Noteworthy in this context is that, compared to the previous studies in this field, German speakers in the present study behaved more “English-like”: they showed overall less bias towards endpoints. Despite testing our German sample group in their native language, the majority have reported being proficient in English and/or Dutch due to their studies, conducted in either English or Dutch as the working language. This means that most of their daily thinking and communication happens in these languages, rather than their native language. Both languages express a progressive aspect viewpoint more frequently compared to German (English in a grammaticalised way, Dutch through lexical means; see also von Stutterheim et al., 2012). Depending on the proficiency and use of other languages (here: Dutch/English used by German speakers) a language-based endpoint bias may have been diminished in our study.

Previous studies of motion cognition have indeed shown that a high degree of exposure to a second language, and immersion in an L2 context, can alter native language patterns in perceiving and speaking about motion events (e.g., Bylund & Jarvis, 2011; Brown & Gullberg, 2010; Athanasopoulos et al., 2015), a process called 'conceptual restructuring' (Jarvis & Pavlenko, 2008). With varying strength depending on many factors, a language user can shift their perspective-taking after their cognitive system gets 'trained' in a new way of construing events. In this case, native speakers of German, originally used to conceptualising events holistically, were exposed on a daily basis to event descriptions centred on the dimension of ongoingness in English (and to a lesser extent, in Dutch, too). This may have led to a shift in event conceptualisation patterns from strong to weak endpoint-bias. Further, in language users with high proficiency in a second/third language, similar to the participants tested here (university students with active L2 use), the specific language of operation in an experimental context mattered for how events were perceived. For example, in a similarity judgment task, the language of instruction used by the experimenter (English or German) drove the extent to which an endpoint-bias could be observed in German L2 users of English (Athanasopoulos et al., 2015). Participants' 'language mode' (Grosjean, 1985) guided attention to either the endpoints or the ongoing phases when perceiving motion events. Here, German native speakers were instructed by a native German speaker experimenter. Nonetheless, their immersion in a second/third language environment could have influenced their behaviour and thinking patterns. From a psycholinguistic viewpoint, our study thus extends previous study on cross-linguistic research on aspect marking. However, more research is needed to study the effects of L2/L3 proficiency and immersion, plus language of operation, in this specific experimental setting in more detail.

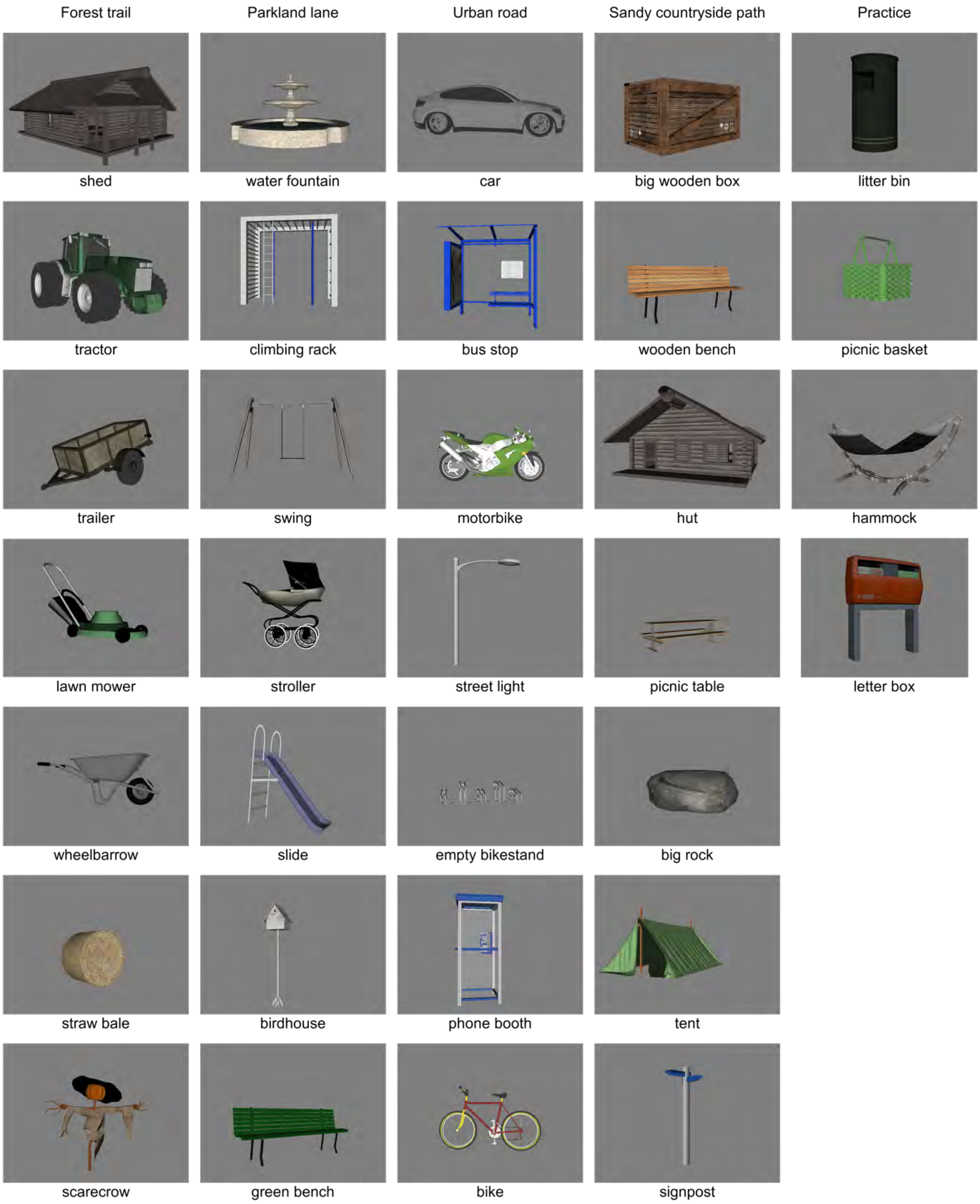
In sum, we observed no cross-linguistic differences in event perception between German and English native speakers, both in a first-person VR set-up and in traditional computer-based control experiments. The finding that looking patterns and control tasks show a similar tendency suggests the VR setup is fit to be used for further psycholinguistic research. In line with this, our participants reported feeling immersed in the setup, and enjoyed moving through the virtual environments. In addition, the setup enabled us to establish looking behaviour over time (i.e., as participants were moving through space). This was something not possible with previous video setups, since participants had less freedom in looking where they desired, but were instead restricted to the scene as it was displayed in the videos. Despite focusing the analysis on specific time-windows, we were able to track looking behaviour and its development over time. From a methodological viewpoint, the immersive VR setting with participants in motion provides a basis for future research avenues by providing a less passive experimental approach, as advocated for also by Hari et al. (2015). The current study is a successful attempt in utilising the CAVE setup in combination with participants in motion, allowing for a first-person immersive experience, and thus more ecologically valid testing of language effects on cognition.

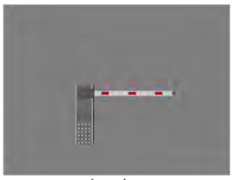
Acknowledgments

We thank Albert Russel and Jeroen Derks for their technical support, Eleanor Huizeling for her helpful discussion of the data, and Sebastian Sauppe for his input concerning growth curve analysis.

5. Appendix

5. 1. Stimuli Items



Forest trail	Parkland lane	Urban road	Sandy countryside path	Practice
 ladder	 parasol	 skateboard	 caravan	
 pitch fork	 shovel	 scooter	 barrier	
 kids' scooter	 tricycle	 bin	 a pile of tree trunks	
 watering can	 bucket	 hose	 sunflower	
 bench	 road blockage	 hydrant	 canoe	

5. 2. Additional Analysis

Results of an additional analysis of Phase 2 using an adjusted (shorter) time-window length of 3.6 seconds, starting at 8.75 seconds (i.e., from when there is a clearly visible divergence between looks to LM and EP objects) until the end of Phase 2 at 12.35 seconds.

Output of the model on EP fixations in Phase 2 (glmer (Hits ~ Language * ($Time^1$ + $Time^2$ + $Time^3$) + Nuisance + (1 | Item-pair) + (1 | Participant))

Fixed Effects	β Estimate	SE	*z*-value
Intercept	-3.706	0.181	-20.489***
Language Eng	-0.009	0.222	-0.04
$Time^1$	2.222	0.2	11.120***
$Time^2$	-0.308	0.2	-1.537
$Time^3$	-0.835	0.2	-4.173***
Nuisance	1.974	0.01	188.820***
Language Eng : $Time^1$	-0.424	0.271	-1.566
Language Eng : $Time^2$	-0.316	0.271	-1.166
Language Eng : $Time^3$	0.326	0.27	1.206

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

CHAPTER 6

SUMMARY AND GENERAL DISCUSSION

The core research questions of the chapters outlined in this thesis revolved around to what extent and under which conditions language guides event perception and memory—using grammatical aspect as a test case.

Theoretical accounts in the fields of event cognition and psycholinguistics indicate that predictive processes underlie event understanding (e.g., Altmann & Mirković, 2009; Radvansky & Zacks, 2014; Ünal, Yi, & Papafragou, 2019). Theoretical approaches from both fields outline analogous dimensions regarding the construction of event (or situation) models, including the spatio-temporal framework within which a model is constructed, and a number of other event dimensions (e.g., protagonists) that are critical to event comprehension. Change along one (or more) event dimension(s) can result in no longer being able to anticipate how the event will unfold. This in turn leads to event model updating, the potential detection of event boundaries (cf. Event Segmentation Theory; EST, Zacks & Swallow, 2007), and commitment of the event to memory. In event cognition theory, language is discussed as a means to experience and learn about events. Through linguistic input, we understand events and their structure. Constructing an event model on the basis of language can support us in mapping out what might happen next and what a given event is about. However, to date, not many studies aimed at specifically manipulating language with a focus on its implications for event cognition theory, pointing towards a gap in the literature.

Events are understood as units of experience with a beginning and endpoint, in which time is seen as a framing pillar of any event. Events can be experienced through language, and indeed any linguistic description of an event provides information about its timing and duration (Madden & Ferretti, 2009). Time in language can guide the construction of an event model by making explicit the event's (temporal) framework, which may have consequences for how we view associated dimensions such as agents and objects, and their relation to one another. Importantly, time can be expressed through different linguistic means, and this thesis specifically studied (grammatical) aspect, which allows for different temporal viewpoints on events. Specifically, the (English) progressive has been found to highlight an event's internal features since it expresses ongoingness. By contrast, the (English) perfective presents an event as holistic, and includes a focus on an event's endpoint (Klein, 1994, 2009; Madden & Ferretti, 2009; Madden & Zwaan, 2003). In line with this view, our hypotheses stated that perfective aspect should activate or guide attention to event entities characterising goal states. The hypotheses were derived from existing work on linguistic experience and event understanding. In behavioural work aspect has been shown to affect the reaction times to pictures showing specific event phases (Madden & Zwaan, 2003), contents of event descriptions (Flecken, Gerwien, Carroll, & von Stutterheim, 2015; von Stutterheim, Andermann, Carroll, Flecken, & Schmiedtová, 2012), mouse-tracking duration and drop-locations (Anderson, Matlock, & Spivey, 2013), and event similarity judgments (Athanasopoulos & Bylund, 2013). Using online measures, effects of aspect have been observed both in eye gaze (Flecken et al., 2015) and ERP patterns (Becker, Ferretti, & Madden-Lombardi, 2013). Overall, studies on linguistic relativity, which contrast verbal to nonverbal tasks, tend to report stronger effects of aspect on event cognition in verbal tasks (Athanasopoulos & Bylund, 2013).

In my experimental chapters, I contrasted the perfective and progressive aspect in English, and studied how aspect usage affects our understanding of and memory for events. The main goal of this thesis was to add to a growing body of research, which suggests that language, and specifically aspect, may affect event model construction and thus guide event comprehension. Methodologically, all studies in this thesis built on existing work in psycholinguistics, yet the interpretation of the results was explicitly embedded in event cognition theory. The broad research questions addressed were (1) to what extent language can activate specific visual representations descriptive of an event, (2) whether such representations can be activated from limited language contexts, (3) how language can affect memory for events and

(4) to what extent previous work on linguistic experience and event cognition generalises to ecologically valid settings. To this end, in Chapters 2-4, I studied change-of-state events, in which an object is irreversibly changed by an action. In Chapter 5, I studied motion events, in which an entity is engaged in a change of location.

1. Summary of the Results

In Chapter 2, I used a sentence-picture verification task adapted for EEG. I presented participants with SVO sentences, where verb semantics suggested a change-of-state in an object (e.g., "*John chopped/was chopping the onion.").* Following the sentence, participants saw a picture of the object in its original unchanged state (NSC), in its changed state (SC), or a picture of an unrelated object (U, control condition). For each picture, they had to indicate whether the object was mentioned in the sentence. EEG was recorded during the presentation of the sentences, and ERPs were analysed from picture onset. For SC and NSC pictures, a P300 was observed. The P300 was larger for SC pictures overall, suggesting SC pictures best matched the sentences. Additionally, SC pictures were processed differently after sentences in perfective compared to progressive aspect. Specifically, the P300 for SC pictures was larger for sentences in the perfective than the progressive. This suggests that the perfective, highlighting the endpoint, was a better match for the SC pictures since it denoted change. The results suggested that both the verbs and the perfective aspect activated specific visual representations of changed objects—a modulating effect of language on event comprehension

Chapter 3 followed up on these results in two web-based experiments measuring reaction times (RTs). In Experiment 1, participants carried out the sentence-picture verification task with the same stimuli as used in Chapter 2. While no effect of Aspect was observed in Experiment 1, the effect for picture type was replicated: SC pictures were overall responded to fastest. This suggests that participants simulated the event during sentence comprehension, which led to the activation of a specific object state. The activated changed state of the object was driven by the semantics of the verbs, which all expressed state change (e.g., chop, peel, slice). In Experiment 2, the object noun was decisively omitted (e.g., "*John chopped/was chopping the...*")*,* to test to what extent people represent state change when the object is unknown. Again, SC pictures were responded to fastest, suggesting state change can be conceptualised independent of the identity of the specific object at hand. An interaction of Picture type and Aspect was also observed: NSC pictures were responded to faster after perfective compared to progressive sentences. The perfective aspect seems to have activated the original state of the object (NSC), whereas the progressive did not (or did so to a lesser extent). In sum, the results from both experiments highlight verb semantics guided event understanding, which is relevant to both full and limited sentence contexts, again supporting the idea that specific linguistic information is taken into account rapidly for event model construction.

Chapter 4 focused on event memory in a web-based study. In two experiments, I looked at whether aspect acts as a cue for memory prior to perception (Experiment 1) and/or as a means for memory encoding after event apprehension (Experiment 2). In Experiment 1, participants read aloud sentences in which aspect was manipulated (e.g., "*She chopped/was chopping an onion*."), after which they watched video clips of the events. Videos either stopped playing before the end of the event (in-progress) or at completion (ceased). In addition, event type was manipulated; half of the video clips showed events with state change (SC) and the other half events without (NSC). Experiment 2 used the same stimuli, but reversed the order such that participants first watched the video clip before reading the accompanying sentence. In a surprise memory task, participants saw video stills and had to indicate whether the still correctly showed how the videos had ended. We analysed memory accuracy, RTs on the correct responses, and signal detection measures d' and C, reflecting sensitivity and response bias, respectively. The results of both experiments highlight an interaction of Event Type and Video Status. Participants were more accurate, faster, and more sensitive in responding to in-progress videos of NSC events (as compared to in-progress SC videos), and in responding to ceased videos of SC events (as compared to ceased NSC videos). In Experiment 2, we further observed an interaction of Aspect and Video Status, such that memory was better for ceased videos presented in the progressive (compared to the perfective) and better for in-progress videos presented in the perfective (compared to the progressive). In sum, the results point to semantics as affecting event memory both for perception as well as for encoding, since memory was somewhat more accurate for completed SC and incomplete NSC events. The effect of Aspect shows a mismatch effect facilitating event memory when language is provided for encoding (after the video) yet not for perception (before the video). The results suggest that effects of language on event memory seem to depend on whether language is provided prior to or after event apprehension.

In Chapter 5, I built on existing cross-linguistic research on motion event perception and tested speakers of an aspectual (English) and a non-aspectual (German) language in an eye-tracking study. Participants engaged in a dynamic non-verbal virtual reality (VR) task, in which they walked on a treadmill across virtual terrains. They encountered objects alongside the road (landmarks) and potential destinations at the end of the road (endpoints). Participant's eye gaze to the objects was recorded to identify a potential endpoint bias that has previously been observed for speakers of non-aspectual languages. After the VR task, participants engaged in an event description task and a similarity judgement task both showing motion events. No cross-linguistic differences were observed in any of the tasks. This indicates that language did not elicit an endpoint bias, and previously observed effects did not replicate to an ecologically valid setting.

First, I will revisit each of the main research questions and highlight the results with regards to the effects of language on event cognition. I will provide possible explanations for the observed patterns across the chapters, especially concerning the findings on aspect. Then, I will integrate all findings to sketch the broader theoretical implications for the role of language on event cognition, before outlining future directions.

2. To What Extent Can Language Activate Specific Visual Representations Descriptive of an Event, and Can Such Representations Be Activated From Limited Language Contexts?

Throughout Chapters 2 and 3, the verbs described an irreversible state change in patient objects. Although verb semantics were not explicitly manipulated, the findings align with psycholinguistic studies showing that linguistic input describing a particular shape or orientation (cf. Zwaan & Pecher, 2012) leads to a specific visual representation to be activated (a 'mental image'). In turn, this leads to facilitated processing for images that show the implied shape/orientation in a sentence-picture verification task. The results in this thesis align with these findings. Specifically, across Chapters 2 and 3, SC pictures following verbs suggesting state change were processed as better matching the sentences compared to NSC pictures that do not evidence change. This was found in RT and ERP measures, with participants showing faster RTs as well as larger P300 responses (indicating match between sentence and picture) for SC pictures. Similar results were also observed in a limited sentence context (Chapter 3, Experiment 2), where RTs were also fastest for SC pictures.

Regarding aspect, the interaction effect in Chapter 2 pointed towards facilitated processing of SC pictures after the perfective compared to the progressive. This finding is in line with previous work suggesting aspect modulates the focus on specific event phases (e.g., Madden & Zwaan, 2003). However, this effect was absent in the behavioural follow-up (Chapter 3, Experiment 1). This is in contrast to behavioural work such as the sentence-picture verification task reported in Madden and Zwaan (2003), who used an analogous setup to the studies reported in this thesis, and do find effects of aspect. They also used verbs suggesting a state change (*accomplishments*, Vendler, 1976), yet their picture stimuli showed in-progress and completed pictures. In their study, participants saw fewer items (*N* = 26 and 28) compared to what was used in Chapter 3, Experiment 1 (*N* = 204). While the aspect manipulation may have stood out in the earlier study, participants in our studies had more of a processing load over time. Consequently, they may have increasingly relied on the verb semantics above aspect to complete the sentence-picture verification task. An interaction with Aspect observed for the limited sentence context (Chapter 3, Experiment 2), suggested facilitated processing for NSC pictures but not for SC pictures. The results indicate that in a limited context participants may make use of all details of the sentence, including aspect, to build their event model.

Overall, the results from Chapters 2 and 3 suggest verbs carry at their core the type of action, which gives insights into likely patient objects and their affordances. The results from the limited sentence context (Chapter 3, Experiment 2) especially highlight, that visual representations of an object’s state can be activated by the verb alone. Additional linguistic information such as object nouns, or indeed aspect, are less critical to mapping out the event. That said, concerning change-of-state events, aspect seems to be taken into account under specific conditions, such as when the processing load allows for it, or when the linguistic context is limited.

3. How Can Language Affect Memory for Events?

While the effects of language on event memory have been studied to an extent (e.g., Loftus & Palmer, 1974; Salomon-Amend, Radvansky, Anderson, Zwaan, & Eerland, 2017; Santín, van Hout, & Flecken, 2021), the literature is sparse. Recently, Zacks (2020) has suggested that event segmentation, which may be modulated by language, is the basis of the event memory formation. In this sense, Chapter 4 contributed to the emerging literature by suggesting new avenues for research on language and event memory. Regarding aspect and event memory, a previous study has shown that event content is more available after the progressive compared to the perfective (Hart & Albarracín, 2009). More recent work by Salomon-Amend et al. (2017) revealed better verbatim memory accuracy after sentences in the perfective compared to the progressive. Both studies used vastly different setups and research foci, also when compared to the experiments outlined in Chapter 4, which complicates integrating this chapter with existing work. The setup in Chapter 4 is most similar to work by Skordos, Bunger, Richards, Selimis, Trueswell and Papafragou (2020), since they also manipulated when language was introduced as a possible tool to memory (i.e., prior or after event apprehension), and since memory was tested in a surprise task. In their study, they found that when language was provided by the experimenter before video clips showing events, or produced by the participants themselves after event video clips, language impaired memory. Skordos et al. (2020), however, did not explicitly manipulate aspect, and also examined motion events rather than action events. They also note that for other types of (event) contexts, language may boost memory. In Chapter 4, no effects of Aspect were observed when language was presented prior to the event videos (Experiment 1), yet there was an effect of Aspect when language was presented after the video (Experiment 2).

In sum, the role of language on memory for events deserves further investigation. While the role of verbs was not explicitly manipulated in Chapter 4, existing research points towards verbs guiding event memory (e.g., Loftus & Palmer, 1974, who manipulated verb semantics; Skordos et al., 2020, who manipulated manner vs path verbs). With regards to aspect, the results add to and highlight the variability between existing studies. As a consequence, the role of aspect in event memory remains unclear and deserves additional attention with a more systematic approach. Overall, the experiments in Chapter 4 highlight that the effects of language seem to be contingent on whether language is provided as a tool for event memory prior or after an event.

4. To What Extent Does Previous Work on Linguistic Experience and Event Cognition Generalise to Ecologically Valid Settings?

In Chapter 5, no effects of aspect were observed. In the VR part of the study, participants engaged in a nonverbal eye-tracking task. Effects of aspect were also absent in the two verbal control tasks (event description and similarity judgement task) following the VR task. This is interesting since effects of aspect—specifically, a goal bias for speakers of non-aspectual languages such as German—have been found across multiple studies using the same setup and stimuli as the control tasks. The absence of an effect of aspect in Chapter 5 may be due to observed effect being overestimated in previous work, and more research may be needed to establish the cross-linguistic effects of aspect on motion event perception in ecologically valid settings. Our sample group showed varying degrees of bilingualism, which may have washed out potential effects of aspect. That said, being bilingual and flexibly switching between languages (which may or may not differ in their aspect marking) is becoming the norm. In that sense, effects of aspect are likely dynamic and may be absent depending on the specific language use patterns of people.

5. The Role of Language on Event Cognition

Overall, the results observed throughout all chapters warrant a discussion of the effects of language on event cognition from two perspectives. As discussed in brief above, effects of (verb) semantics have been observed throughout the thesis. However, the role of aspect on event cognition, which was used as a test case in this thesis, yielded mixed results.

The results, especially those of Chapters 2 and 3, indicate that situation model construction was guided by the specific verbs used in this thesis. The findings contribute to Situation Model theory, and specifically Zwaan, Langston and Graesser's (1995) Event-Indexing Model. The Event-Indexing Model states that events can be captured by the verb in a sentence. During language comprehension, readers monitor the incoming language input to update their situation model and eventually make sense of the described situation. Model updating occurs based on changes in event dimensions, such as time and causality. According to Zwaan et al. (1995), verbs carry information about these event dimensions, and can also capture information about object state change. Verbs could be seen as capturing the essence of a situation or event, because they can denote the type of action and map out the trajectory of change. Through *Aktionsart*, they can also provide information about the event's inner temporal structure. Indeed, empirical work also highlights the importance of verbs in mapping out the event. For example, Hindy, Altmann, Kalenik and Thompson-Shill (2012), and also Kang et al. (2019, 2020), contrasted verbs suggesting state change (e.g., "*drop*") to those that did not (e.g., "*choose*"), which led to different object state representations in people's minds.

Recall that the verbs used in Chapters 2 and 3 were specifically chosen to describe an irreversible change-of-state in an object. This change reflects in a changed object as represented by the SC pictures. As discussed, findings on the SC pictures align with psycholinguistic studies on the shape effect, for which the activation of a specific 'mental image' has been observed (e.g., Zwaan & Pecher, 2012). Zwaan and Pecher (2012) discuss their results in terms of mental simulation theory. While mental simulation theory, and specifically perceptual symbol theory (Barasalou, 1999), were not an explicit focus of this thesis, the obtained results nevertheless complement this theoretical view. Upon reading a state change verb in a sentence context, participants seem to activate the changed state of the object represented by the SC pictures. Linking back to the Event-Indexing Model (Zwaan et al., 1995), these results indicate that an event dimension such as causality in a situation model is tied predominantly to the verb. Verbs carry at their core the type of action, which gives insights into the likely event participants, including patient objects and their affordances.

From the perspective of event cognition theory, change is fundamental to how we understand and memorise events. Change along event dimensions such as time, people and objects, helps us segment ongoing experiences into discrete events (Event Segmentation Theory; EST, Zacks & Swallow, 2007). Altmann and Ekves' (2019) recent theoretical account of event cognition (Intersecting Object Histories; IOH) states that we understand events by tracking the change an object undergoes throughout an event. Upon reading about an event, people indeed construct an event model of the situation. In Chapters 2 and 3, the linguistic descriptions of the events inherently carried information highlighting the affectedness of an object, with the resultant changed state represented by the SC pictures. In line with the results, understanding state change (cf. IOH, Altmann & Ekves, 2019) hinges on the verb, as it specifies the relation between the action and the object. A state change verb allows us to map out what the event is about, how it will unfold, and this in turn affects event understanding and responses to experienced events. Whether the sentence context is complete or limited may play a secondary role in simulating and understanding the essence of a (state change) event as described by the verb.

Overall, the (semantics of the) verbs do seem to have a clear and stable effect on event conceptualisation, which was observed especially in Chapters 2 and 3 of this thesis[13]. This makes sense given that verbs define events (cf. Event-Indexing Model, Zwaan et al., 1995; Altmann & Mirković, 2009): They are central in indicating how events are viewed and conceptualised. Verbs afford the classification of events (cf. Madden & Ferretti, 2009), because they encapsulate information about the nature of the events they describe (e.g., whether they are states, have inherent endpoints etc.). This inherently carries information about other event entities, and how they might relate to one another. For example, in a chopping event, the action necessarily involves an agent and affected object. Linking back to IOH (Altmann & Ekves, 2019), the results throughout this thesis suggest verb semantics describing an object undergoing change are the 'driving force' behind successful event model construction for change-of-state events. In change-of-state events, objects do not seem to be represented in their prototypical instantiations as is common for object labels (cf. Lupyan, 2012). Instead, in context with a verb, objects are represented in accordance with their affordances in conjunction with the verb semantics (Altmann & Kamide, 1999; Kang et al., 2019; see also Altmann & Ekves, 2019 for their discussion on object types, tokens and token-states).

[13]In Chapter 4, verb semantics were not explicitly manipulated, but classification of the event video clips into those with an inherent endpoint (SC) and without (NSC) was arguably supported by the accompanying sentences, which directly contrasted events with and without a state change through the choice of verbs used. As discussed above, previous work has found effects of verb semantics on event memory.

In comparison to verbs, it is less clear how aspect influences event cognition, especially given the results observed in this thesis. Previous work has shown that aspect may affect whether people perceive events as completed or ongoing. Among other things, this directly impacts how we judge event duration (Magliano & Schleich, 2000; Flecken & Gerwien, 2013), and which event features may be foregrounded and remembered (e.g., Carreiras, Carriedo, Alonso, & Fernández,1997; Hart & Albarracín, 2009; Madden & Zwaan, 2003; Madden & Therriault, 2009). Radvanksy and Zacks (2014) acknowledge the potential of aspect in event model construction. Based on empirical work such as the examples above, they suggest that verb aspect 'guides segmentation and model construction' (Radvansky & Zacks, 2014, p. 73). However, the results observed throughout this thesis are in part inconsistent with existing work showing effects of aspect, despite similar experimental setups and stimuli. As discussed above, effects of aspect seem to be highly context-dependent (i.e., contingent on specific task demands and language use patterns), and the nature of the role of aspect in event cognition thus remains somewhat elusive. In other words, although aspect can affect the construction of event models, thereby influencing event understanding and memory, this evidently does not seem to occur always, nor by default.

To sum up, verbs act as a scaffold for the event structure and how it is mentally represented. Given the variability of results regarding aspect—both in the existing literature and in this thesis—, aspectual information seems peripheral in the sense that it does not define the type of event in the same way as verbs do. Rather, aspect expresses how the particular event could be perceived in temporal terms. In other words, verbs define who is doing what to whom, which seems to be more central to event conceptualisation than the temporal perspective conveyed by aspect marking on the verb. Overall, the results of this thesis demonstrate that language—or linguistic experience—can support event model construction as acknowledged by event cognition theory (Radvanksy & Zacks, 2014; Altmann & Ekves, 2019). This affects our understanding of events, as we continue to talk about what we did when, and read about what happens to others (Zacks, 2020). The effects of language on event cognition in this thesis point towards the following: Firstly, verbs do guide event understanding, as they are paramount to the conceptualisation of events (Altmann & Kamide, 1999; Zwaan et al., 1995). This seems to be particularly relevant for change-of-state events (cf. IOH, Altmann & Ekves, 2019), because verbs convey information about whether or not the event involves change, the type of change, and whether change is inflicted upon patient objects. Secondly, and in line with previous empirical work (e.g., Athanasopoulos & Bylund, 2013), effects of language have been observed in this thesis under verbal conditions, in which language was explicitly provided during the tasks. Lastly, aspect may have the potential to affect event cognition, but such effects do not necessarily arise at all times. Given the variability of results in this thesis and in the existing literature, the conditions under which effects of aspect do arise, are difficult to pinpoint. This indicates that interpretations of previous observations of aspect effects may need to take into account, among others, task design, stimuli and language(s) tested. Consequently, this suggests aspect may play a more minor role in shaping event cognition than previously assumed.

6. Future Directions

Given the above, the nature of the role aspect in event cognition may benefit from additional research. In the case of change-of-state events, the sentence-picture verification task reflects a useful paradigm in assessing how aspect may affect (visual) representations of event entities. The outlined studies from Chapters 2 and 3 could be extended by including perfect aspect in addition to perfective and progressive to get a clearer idea as to how different aspectual viewpoints may constrain event model construction. According to Klein's view on time in language (1994, 2009), perfective and perfect should both foreground end states reflected in SC pictures. The perfective may also lead to activation of NSC object states, while the perfect should not. This is because the perfective represents the event holistically, which may include an object's initial state, whereas the perfect mainly highlights the changed state. Studying which object states are activated as a result of aspect, will contextualise the position made in IOH that multiple object states appear to be activated simultaneously. The availability of the initial object state (i.e., the NSC picture) may be subject to aspect modulation, with NSC pictures possibly more available after the perfective compared to the perfect. In addition, lab-based behavioural and EEG versions of the studies outlined in Chapter 3 will highlight specifically under which conditions aspect affects event cognition. Cross-linguistic contrasts would allow for additional insights into the role of verbs, and the potential effects of aspect. English verbs, such as the ones used in this thesis, can express state change (e.g., chop), but do not encode the completion or goal state. By contrast, in Mandarin, serial verb constructions entail event completion (e.g., chop-pieces), and thus state change (Chen, 2018). In this sense, the perfective viewpoint encapsulating the goal state is encoded in such serial verb constructions. Contrasting these two languages would allow to test further whether aspectual viewpoint information can play an important role, if it is part of the central defining part of the event, i.e., the verb.

The experiments in Chapter 4 were driven by the research question of how language can affect event memory, and to this end language was provided after or prior to encoding of the event video clips. In terms of the research question, Chapter 4 is similar to recent work by Skordos et al. (2020), who systematically took into account a variety of ways in which language may influence event memory. For example, they had participants engage in verbal and non-verbal tasks, and also manipulated whether language was presented prior to the event or after the event (by self-production of the participants). Additionally, and in line with suggestions by Skordos et al. (2020), it would be worthwhile to explore further to what extent language may affect memory during the recognition (i.e., the memory test) phase. Since the experiments in Chapter 4 were web-based, it was impractical to control whether participant could use language during the memory task to support their responses, and whether they may have been using language in line with our aspect manipulation. This leaves open the question at what point exactly aspect has played a role in boosting event memory. A lab-based study may provide clearer insights, by allowing for control over whether language is used by the participants, which would enable testing more specifically whether language can affect event memory only during inspection, or also later on, and how long-lasting effects of language on memory may be.

Lastly, concerning the applied implications of aspect marking, studies would benefit from comparing languages that differ in their aspect marking. In the current thesis, English was used as a test case, since the majority of work has been carried out in that language making comparisons to existing work more straightforward. However, note that the perfective in English is conveyed through the simple past tense (as used in this thesis, e.g., "*chopped*"), and is thus unmarked. As a result, (past) tense and (perfective) aspect are conflated in English. This is in contrast to the progressive, which is marked (e.g., "*was chopping*"). As a result, some suggest that the English simple past may actually be aspectually neutral (Schmiedtová & Flecken, 2008), such that it can be flexible and adapt to convey a non-perfective meaning as is the case, for example, for habitual expressions (in the past) (e.g., "*She took the bus to work.*"). The progressive, too, is flexible in the sense that it can include events' endpoints, especially when the verb semantics drive an endpoint-biased event model. When used with verbs suggesting events with an inherent endpoint (telic verbs), such as the ones used in this thesis, the simple past in English is accepted to reflect a perfective viewpoint on an event (Klein, 1994). In this sense, English aspect can be considered somewhat underspecified and aspectual meaning in

English is influenced by additional context, such as verb semantics. Future research should thus aim to extend the work outlined in this thesis to include languages with a binary aspectual opposition between progressive and perfective, and in which aspect is not conflated with tense (e.g., Russian), or in which aspect is not marked but (lexically) frequently used (e.g., Dutch). This will offer clear insights into whether effects of (grammatical) aspect on event model construction indeed derive from aspect.

In addition to cross-linguistic work, taking into consideration culture may also open up interesting new research avenues: Recent work by Swallow and Wang (2020) has shown that cultural background may affect event cognition; using an event segmentation task, they found that US and Indian participants segmented video clips of the same actions (e.g., making coffee; videos were filmed in the US and India) differently, reflecting cultural preferences for visual and goal-related changes. Culture and language may of course interact. A triangulation design, in which two of three sample groups share the same language (or grammatical specifics, such as aspect), and two share the same culture, may offer a useful approach into studying which effects on event cognition are culture-specific and which are language-specific.

7. Concluding Remarks

The role of language in cognition has been debated for decades. An extreme Whorfian stance on the language and thought debate has been abandoned, and research in this domain is taking more nuanced approaches. Perpetuating binary examinations of whether or not effects of language on cognition do arise, or whether or not language is separate from cognition, does not provide meaningful insights into the human experience. Instead, empirical work must emphasise how and when language can affect cognition, and identify domains for which language can guide cognitive processing (see Gentner, 2016). Event cognition appears to be one of those domains, and this thesis outlined empirical research in the interface of psycholinguistics and event cognition by drawing on existing work in both fields. Embedding the interpretations of results in both fields widens their scope and invites future studies to follow suit. 'Events are the stuff of all our lives.' (Zacks & Tversky, 2001, p. 19), and I hope this thesis adds to our understanding of what this 'stuff' entails, while also providing a basis for future research.

To inspire future studies in this domain, the materials from all chapters are available via the MPI Language Archive.

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. *Psychophysiology, 56*(12), e13451.

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.

Altmann, G. T., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review, 126*(6), 817-840.

Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing *Cognitive Science, 33*(4), 583-609.

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

Amoruso, L., Gelormini, C., Aboitiz, F. A., González, A., Manes, F., Cardona, J., & Ibanez, A. (2013). N400 ERPs for actions: building meaning in context. *Frontiers in Human Neuroscience, 7*, 57.

Anderson, S., Matlock, T., & Spivey, M. (2013). Grammatical aspect and temporal distance in motion descriptions. *Frontiers in Psychology, 4*, 337.

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.

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N., & Thierry, G. (2015). Two languages, two minds: flexible cognitive processing driven by language of operation. *Psychological Science, 26*(4), 518-526.

Athanasopoulos, P., Damjanovic, L., Burnand, J., & Bylund, E. (2015). Learning to think in a second language: Effects of proficiency and length of exposure in English learners of German. *The Modern Language Journal, 99*(S1), 138-153.

Attneave, F. (1957). Transfer of experience with a class-schema to identification-learning of patterns and shapes. *Journal of Experimental Psychology, 54*(2), 81–88.

Baggio, G., Van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language, 59*(1), 36-53.

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

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577-660.

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.

Blanco-Elorrieta, E., & Pylkkänen, L. (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences, 22*(12), 1117-1126.

Brown, A. & Gullberg, M. (2010). Changes in encoding of path of motion after acquisition of a second language. *Cognitive Linguistics, 21*(2). 263–286.

Bylund, E. & Jarvis, S. (2011). L2 effects on L1 event conceptualization. *Bilingualism: Language and Cognition, 14*(1), 47-59.

Carreiras, M., Carriedo, N., Alonso, M. A., & Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Memory & Cognition*, *25*(4), 438-446.

Cañigueral, R., & Hamilton, A. F. D. C. (2019). The role of eye gaze during natural social interactions in typical and autistic people. *Frontiers in Psychology, 10*, 560.

Chen, J. (2018). "He killed a chicken, but it didn't die": An empirical study of the lexicalization of state change in Mandarin monomorphemic verbs. *Chinese Language and Discourse*, *9*(2), 136-161.

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems.* Cambridge University Press.

Connell, L. (2007). Representing object colour in language comprehension. *Cognition, 102*(3), 476-485.

Cruz-Neira, C., Leigh, J., Papka, M., Barnes, C., Cohen, S. M., Das, S., ... & Vasilakis, C. (1993). Scientists in wonderland: A report on visualization applications in the CAVE virtual reality environment. In *Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium* (pp. 59-66). IEEE.

Curiel, J. M., & Radvansky, G. A. (2014). Spatial and character situation model updating. *Journal of Cognitive Psychology, 26*(2), 205–212

Eichert, N., Peeters, D., & Hagoort, P. (2018). Language-driven anticipatory eye movements in virtual reality. *Behavior Research Methods, 50*(3), 1102-1115.

Feller, D. P., Eerland, A., Ferretti, T. R., & Magliano, J. P. (2019). Aspect and narrative event segmentation. *Collabra: Psychology, 5*(1), 12.

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.

Flecken, M. (2010). *Event conceptualization in language production of early bilinguals*. LOT dissertation series 256. Utrecht: Igitur.

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

Flecken, M., Gerwien, J., Carroll, M., & von Stutterheim, C. (2015). Analyzing gaze allocation during language planning: a cross-linguistic study on dynamic events. *Language and Cognition, 7*(1), 138-166.

Flecken, M., Walbert, K., & Dijkstra, T. (2015). 'Right now, Sophie∗ swims in the pool?!': Brain potentials of grammatical aspect processing. *Frontiers in Psychology*, *6*, 1764.

Flecken, M., von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Findings from a cross-linguistic non-verbal recognition task. *Language and Cognition, 6*(1), 45-78.

Gentner, D. (2016). Language as cognitive tool kit: How language supports relational thought. *American Psychologist*, *71*(8), 650-657.

Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. *Cognition*, *180*, 225-237.

Goldschmied, N., Sheptock, M., Kim, K., & Galily, Y. (2017). Appraising Loftus and Palmer (1974) post-event information versus concurrent commentary in the context of sport. *Quarterly Journal of Experimental Psychology, 70*(11), 2347-2356.

Hari, R., Henriksson, L., Malinen, S., & Parkkonen, L. (2015). Centrality of social interaction in human brain function. *Neuron, 88*(1), 181-193.

Hart, W., & Albarracín, D. (2009). What I was doing versus what I did: Verb aspect influences memory and future actions. *Psychological Science, 20*(2), 238-244.

Heyselaar, E., Segaert, K., Wester, A. J., Kessels, R. P., & Hagoort, P. (2015). Syntactic operations rely on implicit memory: Evidence from patients with amnesia. In *The Individual Differences in Language Processing across the Adult Life Span Workshop.*

Heyselaar, E., Peeters, D., & Hagoort, P. (2021). Do we predict upcoming speech content in naturalistic environments? *Language, Cognition and Neuroscience, 36*(4), 440-461.

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.

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.

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.

Hong, J. P., Ferretti, T. R., Craven, R., & Hepburn, R. D. (2019). Time, verbs, and imagining events. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 73*(1), 55-63.

Huff, M., Meitz, T. G., & Papenmeier, F. (2014). Changes in situation models modulate processes of event perception in audiovisual narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1377-1388.

Jarvis, S. & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition.* Routledge.

Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., & Altmann, G. (2019). The influence of state change on object representations in language comprehension. *Memory & Cognition, 48*(3), 390-399.

Kang, X., Joergensen, G. H., & Altmann, G. T. (2020). The activation of object-state representations during online language comprehension. *Acta Psychologica*, *210*, 103162.

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.

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

Klein, W. (2009). How time is encoded. In W. Klein & P. Li (Eds.), *The expression of time* (pp. 39-82). Mouton de Gruyter.

Knoeferle, P. (2015). Visually situated language comprehension in children and in adults. In R. Mishra, N. Srinivasan & F. Huettig (Eds.), *Attention and vision in language processing* (pp. 57-75). Springer.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*(3), 557-577.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

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.

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.

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition, 96*(1), 1-33.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764-766.

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

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Psychology, 3*, 54.

Lupyan, G., Abdel Rahman, R., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in Cognitive Sciences, 24*(11), 930-944.

Madden, C. J., & Ferretti, T. R. (2009). Verb aspect and the mental representation of situations. In W. Klein & P. Li (Eds.), *The expression of time* (pp. 217-240). De Gruyter Mouton.

Madden, C. J., & Therriault, D. J. (2009). Short article: Verb aspect and perceptual simulations. *Quarterly Journal of Experimental Psychology*, *62*(7), 1294-1303.

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition, 31*(5), 663-672.

Madden-Lombardi, C., Dominey, P. F., & Ventre-Dominey, J. (2017). Grammatical verb aspect and event roles in sentence processing. *PloS one, 12*(12), e0189919.

Magliano, J. P., & Radvansky, G. A. (2001). Goal coordination in narrative comprehensions. *Psychonomic Bulletin & Review, 8*(2), 372-376.

Magliano, J. P., Radvansky, G. A., Forsythe, J. C., & Copeland, D. E. (2014). Event segmentation during first-person continuous events. *Journal of Cognitive Psychology, 26*(6), 649-661.

Magliano, J. P., & Schleich, M. C. (2000). Verb aspect and situation models. *Discourse Processes*, *29*(2), 83-112.

Makowski, (2018). The psycho Package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software, 3*(22), 470.

Matlock, T. (2011). The conceptual motivation of aspect. In K.-U. Panther & G. Radden (Eds.), *Motivation in Grammar and the Lexicon* (pp.133-147).

Mirman, D. (2016). *Growth curve analysis and visualization using R.* CRC press.

Misersky, J., Slivac, K., Hagoort, P., & Flecken, M. (2021). The state of the onion: Grammatical aspect modulates object representation during event comprehension. *Cognition, 214*, 104744.

Montero-Melis, G., Jaeger, T. F., & Bylund, E. (2016). Thinking is modulated by recent linguistic experience: Second language priming affects perceived event similarity. *Language Learning*, *66*(3), 636-665.

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.

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.

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.

Pan, X., & Hamilton, A. F. D. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology, 109*(3), 395-417.

Peelle, J. E., & Van Engen, K. J. (2021). Time stand still: Effects of temporal window selection on eye tracking analysis. *Collabra: Psychology*, *7*(1), 25961.

Peeters, D. (2018). A standardized set of 3-D objects for virtual reality research and applications *Behavior Research Methods, 50*(3), 1047-1054.

Peeters, D., & Dijkstra, T. (2018). Sustained inhibition of the native language in bilingual language production: A virtual reality approach. *Bilingualism: Language and Cognition, 21*(5), 1035-1061.

Peeters, D. (2019). Virtual reality: A game-changing method for the language sciences. *Psychonomic Bulletin & Review, 26*(3), 894-900.

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

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition.* 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.

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), e12708.

Salomon-Amend, M., Radvansky, G. A., Anderson, S., Zwaan, R., & Eerland, A. (2017). What's done is done: Verb aspect and verbatim memory. *Collabra: Psychology, 3*(1), 22.

Santín, M., van Hout, A., & Flecken, M. (2021). Event endings in memory and language. *Language, Cognition and Neuroscience, 36*(5), 625-648.

Sauppe, S. (2017). Word order and voice influence the timing of verb planning in German sentence production. *Frontiers in Psychology, 8*, 1648.

Sauppe, S., & Flecken, M. (2021). Speaking for seeing: Sentence structure guides visual event apprehension. *Cognition*, *206*, 104516.

Sassenhagen, J., & Alday, P. M. (2016). A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language, 162*, 42-45.

Schmiedtová, B., & Flecken, M. (2008). Aspectual concepts across languages: Some considerations for second language learning. In *Cognitive approaches to pedagogical grammar* (pp. 357-384). De Gruyter Mouton.

Skordos, D., Bunger, A., Richards, C., Selimis, S., Trueswell, J., & Papafragou, A. (2020). Motion verbs and memory for motion events. *Cognitive Neuropsychology, 37*(5-6), 254-270.

Smith, C. S. (1991). *The parameter of aspect*. Springer.

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.

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.

Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition*, *121*(3), 409-415.

Swallow, K. M., & Wang, Q. (2020). Culture influences how people divide continuous sensory experience into events. *Cognition, 205*, 104450.

Tromp, J., Peeters, D., Meyer, A. S., & Hagoort, P. (2018). The combined use of virtual reality and EEG to study language processing in naturalistic environments. *Behavior Research Methods, 50*(2), 862-869.

Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language, 63*(1), 64-82.

Ünal, E., Ji, Y., & Papafragou, A. (2021). From event representation to linguistic meaning. *Topics in Cognitive Science, 13*(1), 224-242.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* Academic Press.

Vendler, Z. (1976). *Linguistics in philosophy*. Cornell University Press.

Von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics, 50*(4), 833-867.

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*(3), 165-168.

Willems, R. M. (2015). *Cognitive neuroscience of natural language use*. Cambridge University Press.

Withers, P. (2017). Frinex: Framework for Interactive Experiments. doi: 10.5281/zenodo.3522911

Yap, F. H., Chu, P. C. K., Yiu, E. S. M., Wong, S. F., Kwan, S. W. M., Matthews, S., ... & Shirai, Y. (2009). Aspectual asymmetries in the mental representation of events: Role of lexical and grammatical aspect. *Memory & Cognition*, *37*(5), 587-595.

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.

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current directions in Psychological Science, 16*(2), 80-84.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3-21.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29-58.

Zacks, J. M. (2020). Event perception and memory. *Annual Review of Psychology, 71*, 165-191.

Zwaan, R. A. (2008). Time in language, situation models, and mental simulations. *Language Learning, 58*, 13-26.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*(5), 292-297.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PloS one, 7*(12), e51382.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168-171.

Nederlandse samenvatting

Van 's morgens douchen tot onze eerste koffie op het werk tot het lezen van een boek voor het slapengaan - dagelijks maken wij een groot aantal gebeurtenissen mee. Eerdere vergelijkbare ervaringen helpen ons te anticiperen op hoe een nieuwe ervaring zich zal ontvouwen. Cruciaal in dit proces zijn de gebeurteniskenmerken ruimte (*waar* gebeurt het?) en tijd (*wanneer?*), die elke gebeurtenis omkaderen en er structuur aan geven. Er wordt ook rekening gehouden met mensen en voorwerpen (*wie en wat?*). Wanneer deze kenmerken veranderen - bijvoorbeeld wanneer iets van locatie verandert, of een voorwerp in de loop van de tijd verandert - kunnen we misschien niet meer anticiperen op wat er daarna zal gebeuren. We 'chunken' de lopende ervaring tijdens een proces dat segmentatie wordt genoemd. Vervolgens slaan wij de ervaring in ons geheugen op als een specifieke gebeurtenis met een begin en een eind. Het wordt een *event model*, waarin relevante kenmerken van de gebeurtenis, zoals voorwerpen, ruimte en tijd, worden weergegeven. Een event model zal steeds opnieuw worden gebruikt wanneer we iets nieuws ervaren dat lijkt op de opgeslagen gebeurtenis.

We leren over gebeurtenissen door onze directe ervaring ermee, maar ook indirect, door taal. Of we nu het nieuws lezen of een vriend een verhaal horen vertellen, taal kan bepalen hoe we gebeurtenissen begrijpen en onthouden. In deze scriptie heb ik me geconcentreerd op de gebeurtenisdimensie van tijd, die centraal staat in de structuur van een gebeurtenis. Tijd kan op verschillende manieren in taal worden uitgedrukt, en elke beschrijving van een gebeurtenis geeft noodzakelijkerwijs ook informatie over de tijd en/of duur van een gebeurtenis. Ik heb me in het bijzonder gericht op grammaticaal aspect, dat verschillende gezichtspunten op een situatie of gebeurtenis kan bieden. Simpel gezegd stelt het grammaticale aspect (in het Engels) ons in staat om een gebeurtenis te beschrijven als voltooid (perfectief aspect, "*She chopped the onion.*") of aan de gang (progressief aspect, "*She was chopping the onion*."). Eerder onderzoek in de psycholinguïstiek suggereert dat aspect kan beïnvloeden of we een gebeurtenis als wel of niet als beëindigd interpreteren. Dit kan op zijn beurt weer invloed hebben op hoe we gebeurteniskenmerken, zoals voorwerpen in gebeurtenissen, representeren, bijhouden en onthouden.

Eerst onderzocht ik hoe gebeurtenisbeschrijvingen van invloed kunnen zijn op hoe we denken over voorwerpen in gebeurtenissen. Deelnemers lazen zinnen waarin het werkwoord suggereerde dat een voorwerp werd veranderd. De zinnen waren ofwel in de perfectieve ("*John chopped the onion*.") of de progressieve vorm ("*John was chopping the onion*."). Na de zin zagen de deelnemers een afbeelding van het voorwerp in zijn oorspronkelijke, onveranderde staat (NSC) of in zijn veranderde staat (SC), of een afbeelding van een niet-gerelateerd voorwerp (U). Uit de hersenpotentialen van de deelnemers bleek dat plaatjes die de veranderde toestand (SC) toonden, het best bij de zinnen pasten. Wanneer de zinnen in het perfectieve aspect stonden, waarin het einde van de gebeurtenis werd benadrukt, werd dit als een betere overeenkomst met de SC-plaatjes waargenomen dan voor progressieve zinnen. In een vervolgexperiment voerden de deelnemers dezelfde taak uit, maar dan met meting van de reactietijden. De veranderde voorwerpplaatjes (SC) werden over het algemeen als beter passend gezien, wat bleek uit het feit dat er het snelst op werd gereageerd. Deze keer werd echter geen effect van aspect geregistreerd. Met dezelfde zinnen onderzocht ik vervolgens of een gebeurtenisbeschrijving mensen aan het denken kan zetten over de toestandsverandering van een voorwerp, zelfs als het voorwerp onbekend is. Om dit te doen, werd het voorwerp weggelaten uit de zin (bijv. "*John chopped/was chopping the …*"), die voorafging aan de afbeelding. Opnieuw werd het snelst gereageerd op de plaatjes met de veranderde toestand (SC). Bovendien werd er sneller gereageerd op plaatjes in de oorspronkelijke staat (NSC) na perfectieve dan na progressieve zinnen. Het perfectieve aspect lijkt de oorspronkelijke toestand van het voorwerp (NSC) te hebben geactiveerd, terwijl de progressief dat niet deed. Al met al suggereren deze resultaten dat mensen zich tijdens het lezen daadwerkelijk een gebeurtenis in hun hoofd voorstellen, waarbij verandering van toestand wordt geactiveerd - zelfs onafhankelijk van een specifiek voorwerp.

Vervolgens was ik geïnteresseerd in hoe gebeurtenisbeschrijvingen het geheugen kunnen beïnvloeden. In Experiment 1 lazen de deelnemers zinnen waarin het aspect ofwel perfectief was (bijv. "*She chopped an onion*.") of progressief ("*She was chopping an onion*."), en vervolgens keken ze naar videoclips van de gebeurtenissen. In Experiment 2 keken de deelnemers eerst naar de video en lazen daarna de zinnen. De video's stopten met spelen voor het einde van de gebeurtenis (in-progress) of na afronding (ceased). Bovendien werd het type gebeurtenis gemanipuleerd; de helft van de gebeurtenissen toonde een voorwerp dat werd veranderd (state-change, SC), terwijl de andere helft een onveranderd voorwerp toonde (no state-change, NSC). In een verrassingsgeheugentaak zagen de deelnemers plaatjes uit de video's en werd hen gevraagd aan te geven of deze correct weergaven hoe de video's waren afgelopen. Over het algemeen reageerden deelnemers het snelst op gebeurtenissen waarbij het voorwerp onveranderd bleef en de video's stopten voordat de gebeurtenis was afgelopen, en op gebeurtenissen waarbij het voorwerp veranderde en de video stopte wanneer de gebeurtenis was afgelopen. Bovendien was in Experiment 2 de herinnering beter voor video's die gestopt waren en gepresenteerd werden in de progressieve vorm ("*was chopping*") en beter voor video's die bezig waren en gepresenteerd werden in de perfectieve vorm ("*chopped*"). Dit mismatch effect van het aspect werd alleen waargenomen wanneer taal werd aangeboden ná de video, maar niet vóór de video. De effecten van taal op het geheugen van gebeurtenissen lijken dus af te hangen van de vraag of de taal voor of na een gebeurtenis wordt aangeboden.

Tenslotte richtte ik me op bewegingsgebeurtenissen. Eerdere studies naar bewegingsgebeurtenissen hebben aangetoond dat sprekers van niet-aspectuele talen meer aandacht hebben voor (potentiële) bestemmingen in vergelijking met sprekers van aspectuele talen. Ik onderzocht de perceptie van gebeurtenissen bij sprekers van een aspectuele (Engels) en een niet-aspectuele (Duits) taal, die op een loopband liepen over virtuele terreinen die beweging in het echte leven simuleerden. In deze virtual reality (VR) taak kwamen ze voorwerpen tegen, zowel langs de weg als potentiële eindpunten (of bestemmingen). Door VR te gebruiken kon ik onze perceptie van bewegingsgebeurtenissen bestuderen in een omgeving die het echte leven imiteerde. De blik van de deelnemers werd geregistreerd om een mogelijke bias met betrekking tot de eindpuntvoorwerpen vast te stellen, aangezien een dergelijke bias eerder is waargenomen bij sprekers van niet-aspectuele talen. Na de VR taak voerden de deelnemers een beschrijvingstaak uit en een *similarity judgment task*, beide met videoclips van bewegingsgebeurtenissen. Bij geen van de taken werden verschillen tussen de twee taalgroepen waargenomen. Taalachtergrond leidde niet tot een eindpunt-bias. Dit betekent dat eerder waargenomen resultaten zich niet herhaalden in een setting vergelijkbaar met wat we in het echte leven tegenkomen.

Samenvattend tonen de resultaten van dit proefschrift dat taal het begrip van gebeurtenissen en het geheugen kan beïnvloeden. Mensen gebruiken vooral de informatie van het werkwoord (bijv. *"to chop"*) om een event model op te bouwen dat de kern van de gebeurtenis weergeeft. Het werkwoord definieert wie wat doet en met wie, en dit beïnvloedt hoe we gebeurteniskenmerken zoals voorwerpen interpreteren, en hoe deze eruit zien tijdens een zich ontvouwende gebeurtenis. Als we echter kijken naar grammaticaal aspect, zijn de resultaten niet eenduidig. Hoewel aspect van invloed kan zijn op de manier waarop we gebeurtenissen begrijpen en onthouden, treden dergelijke effecten niet noodzakelijkerwijs altijd op.

English summary

From taking a shower in the morning to having our first coffee at work to reading a book before bed – we experience a multitude of events on a daily basis. Previous similar experiences help us anticipate how an ongoing experience may unfold. Crucial to this process are the event features of space (*where* is it happening?) and time (*when?*), which frame any event and give it structure. People and objects (*who and what?*) are taken into account, too. When there is a change in these features – for example, when there is a shift in location, or an object changing over time – we might no longer be able to anticipate what will happen next. We 'chunk' the ongoing experience during a process called segmentation. We then store the experience as a distinct event with a beginning and an end in our memory. It becomes an event model, in which relevant features of the event, such as objects, space and time are represented. An event model will be used again and again when we find ourselves experiencing something new that is similar to the stored event.

We learn about events through our direct experience with them, but also indirectly, through language. Whether we read the news or hear a friend tell a story, language can guide how we understand and remember events. In this thesis I have focused on the event dimension of time, which is central to an event's structure. Time can be expressed in language through different means, and any description of an event necessarily also provides information about an event's time and/or duration. Specifically, I have focused on grammatical aspect, which can offer different viewpoints on a situation or event. Put simply, grammatical aspect allows us to describe an event either as completed (perfective aspect, "*She chopped an onion.*") or ongoing (progressive aspect, "*She was chopping an onion.*"). Previous research in psycholinguistics suggests that aspect may have an impact on whether we understand an event as having ended or not. This in turn may affect how we represent, attend to, and memorise event features, such as objects in events.

First, I explored the question of how event descriptions can affect how we think about objects in events. Participants read sentences where the verb suggested that an object was being changed. Sentences were either in the perfective ("*John chopped the onion.*") or the progressive ("*John was chopping the onion.*"). Following the sentence, participants saw a picture of the object in its original, unchanged state (NSC) or in its changed state (SC), or a picture of an unrelated object (U). Participants' brain potentials revealed that pictures showing the changed state (SC) were perceived to best match the sentences. When sentences were in the perfective, highlighting the event's end, this was perceived a better match for the SC pictures compared to the progressive sentences. In a follow-up experiment, participants carried out the same task, but with reaction times being measured. The changed object pictures (SC) were overall perceived as a better match, evidenced by being responded to fastest. However, this time no effect of aspect was observed. Using the same sentences, I then went on to study whether an event description can make people think about an object's state change even when the object is unknown. To do this, the object was left out from the sentence (e.g., "*John chopped/was chopping the...*"), which preceded the picture. Again, the changed state (SC) pictures were responded to fastest. In addition, original state (NSC) pictures were responded to faster after perfective compared to progressive sentences. The perfective aspect seems to have activated the original state of the object (NSC), whereas the progressive did not. Taken together, these findings suggest that, during reading, people really imagine an event unfolding in their minds, with state change being activated – even independently of a specific object.

Next, I was interested in how event descriptions can affect event memory. In Experiment 1, participants read sentences in which aspect was either perfective (e.g., "*She chopped an* *onion.*") or progressive *("She was chopping an onion.")*, and then watched video clips of the events. In Experiment 2, participants first watched the video and then read the sentences. Videos either stopped playing before the end of the event (in-progress) or after completion (ceased). In addition, the Event Type was manipulated; half of the events showed an object being changed (state-change, SC), while the other half featured an unchanged object (no state-change, NSC). In a surprise memory task, participants saw video stills and were asked to indicate whether it correctly showed how the videos had ended. Overall, participants tended to respond fastest to events where the object remained unchanged and the videos stopped before the event ended, and to events where the object changed and the video stopped when the event had ceased. Additionally, in Experiment 2, memory was better for ceased videos presented in the progressive ("*was chopping*") and better for in-progress videos presented in the perfective ("*chopped*"). This mismatch effect of Aspect was only observed when language was provided after the video yet not for before the video. So, effects of language on event memory seem to depend on whether language is provided prior to or after an event.

Finally, I turned to motion events. Previous studies on motion events have shown that speakers of non-aspectual languages tend to focus more on (potential) goals compared to speakers of aspectual languages. I explored event perception of speakers of an aspectual (English) and a non-aspectual (German) language, who walked on a treadmill across virtual terrains simulating real-life movement. In this virtual reality (VR) task, they encountered objects, both alongside the road and as potential endpoints (or destinations). Using VR allowed me to study our perception of motion events in a setting that mimicked real life. The participants' eye gaze was recorded to identify a potential bias regarding the endpoint objects, as such a bias has previously been observed for speakers of non-aspectual languages. After the VR task, participants engaged in an event description task and a similarity judgement task, both involving video clips of motion events. No differences between the two language groups were observed in any of the tasks. Language background did not elicit an endpoint bias, since previously observed effects did not replicate in a setting similar to what we encounter in real life.

Overall, the results of this thesis highlight that language *can* guide event understanding and memory. People especially use the information carried by the verb (e.g., to chop) to build an event model which captures the essence of what the event is about. The verb defines who is doing what to whom and this impacts how we understand event features such as objects, and what they look like throughout an unfolding event. Looking at grammatical aspect, however, the results are mixed. While aspect may have the potential to affect how we understand and memorise events, such effects do not necessarily arise at all times.

Deutsche Zusammenfassung

Von der morgendlichen Dusche über den ersten Kaffee bei der Arbeit bis hin zum Lesen eines Buches vor dem Schlafengehen – wir erleben jeden Tag eine Vielzahl von Ereignissen. Frühere ähnliche Erfahrungen helfen uns dabei, zu verstehen, wie sich ein neues Erlebnis abspielen wird. Entscheidend für diesen Prozess sind die Ereignismerkmale Raum (wo findet es statt?) und Zeit (wann?), die jedes Ereignis strukturieren. Auch Personen und Gegenstände (wer und was?) werden hierbei berücksichtigt. Wenn sich diese Merkmale ändern –zum Beispiel, wenn etwas seinen Standort wechselt oder sich ein Gegenstand im Laufe der Zeit verändert– können wir möglicherweise nicht vorhersehen, was als Nächstes passieren wird. In einem Prozess, der Segmentierung genannt wird, unterteilen wir das aktuell Erlebte. Wir speichern das Erlebnis dann in unserem Gedächtnis als ein bestimmtes Ereignis (oder Event) mit einem Anfang und einem Ende. Es wird zu einem *event model*, in dem relevante Merkmale wie Gegenstände, Raum und Zeit repräsentiert werden. Ein *event model* wird immer wieder verwendet, wenn wir etwas Neues erleben, das dem gespeicherten Ereignis ähnelt.

Wir lernen über Ereignisse, indem wir sie direkt erleben, aber auch indirekt, durch Sprache. Ob wir nun die Nachrichten lesen oder einen Freund eine Geschichte erzählen hören, Sprache kann uns dabei helfen, Ereignisse zu verstehen und uns an sie erinnern. In dieser Doktorarbeit habe ich mich auf das Ereignismerkmal Zeit konzentriert, das für die Struktur eines Ereignisses von zentraler Bedeutung ist. Zeit kann in der Sprache auf unterschiedliche Weise ausgedrückt werden, und jede Beschreibung eines Ereignisses enthält notwendigerweise Informationen über seine Zeit und/oder Dauer. Ich habe mich insbesondere auf den grammatikalischen Aspekt konzentriert, der verschiedene Blickwinkel auf eine Situation oder ein Ereignis bieten kann. Einfach ausgedrückt, erlaubt uns der grammatikalische Aspekt (im Englischen), ein Ereignis als abgeschlossen (perfektiver Aspekt, "*She chopped the onion.*") oder im Verlauf (progressiver Aspekt, "*She was chopping the onion.*") zu beschreiben. Frühere Forschungen in der Psycholinguistik deuten darauf hin, dass Aspekt einen Einfluss darauf haben kann, ob wir ein Ereignis als abgeschlossen interpretieren oder nicht. Dies wiederum kann sich darauf auswirken, wie wir Merkmale von Ereignissen, wie zum Beispiel Gegenstände, repräsentieren, wahrnehmen und uns einprägen.

Zunächst habe ich untersucht, wie sich Ereignisbeschreibungen darauf auswirken können, wie wir über Gegenstände in Ereignissen denken. Die Versuchspersonen lasen Sätze, in denen das Verb andeutete, dass ein Gegenstand verändert wurde. Die Sätze waren entweder im perfektiven Aspekt ("*John chopped the onion.*“) oder im progressiven Aspekt ("*John was chopping the onion.*“). Nach dem Satz sahen die Versuchspersonen ein Bild des Gegenstandes in seinem ursprünglichen, unveränderten Zustand (NSC) oder in seinem veränderten Zustand (SC), oder ein Bild eines unrelatierten Gegenstandes (U). Die Gehirnpotenziale der Versuchspersonen zeigten, dass Bilder, die den veränderten Zustand (SC) zeigten, als am besten zu den Sätzen passend wahrgenommen wurden. Wenn die Sätze im perfektiven Aspekt standen, der das Ende des Ereignisses hervorhob, wurde dies als bessere Übereinstimmung mit den SC-Bildern empfunden als bei progressiven Sätzen. In einem Folgeexperiment führten die Versuchspersonen dieselbe Aufgabe durch, wobei jedoch Reaktionszeiten gemessen wurden. Auch hier wurden die SC-Bilder im Allgemeinen als besser passend empfunden, was sich darin zeigte, dass auf sie am schnellsten reagiert wurde. Dieses Mal wurde jedoch keine Auswirkung des Aspekts festgestellt. Mit denselben Sätzen untersuchte ich dann, ob eine Ereignisbeschreibung dazu führen kann, dass wir über die Zustandsänderung eines Gegenstands nachdenken, selbst wenn der Gegenstand unbekannt ist. Dazu wurde der Gegenstand aus dem Satz (z. B. "*John chopped/was chopping the*...“) weggelassen. Auch hier wurden auf die Bilder mit dem veränderten Zustand (SC) am schnellsten reagiert. Außerdem wurde auf Bilder im Originalzustand (NSC) nach perfektiven Sätzen schneller reagiert als nach progressiven Sätzen. Der perfektive Aspekt scheint den ursprünglichen Zustand des Gegenstands (NSC) aktiviert zu haben, während der progressive Aspekt dies nicht tat. Alles in allem deuten diese Ergebnisse darauf hin, dass wir uns beim Lesen tatsächlich ein Ereignis vorstellen bei dem eine Zustandsänderung aktiviert wird – auch unabhängig von einem bestimmten Gegenstand.

Als Nächstes interessierte mich, wie Ereignisbeschreibungen das Gedächtnis beeinflussen können. In Experiment 1 lasen die Versuchspersonen Sätze, in denen der Aspekt entweder perfektiv (*"She chopped an onion."*) oder progressiv ("*She was chopping an onion.*") war, und sahen sich dann Videoclips der Ereignisse an. In Experiment 2 sahen die Versuchspersonen zunächst das Video und lasen danach die Sätze. Die Videos wurden entweder bis vor dem Ende des Ereignisses (im Verlauf) oder nach dessen Abschluss (abgeschlossen) abgespielt. Außerdem wurde die Art des Ereignisses manipuliert; die Hälfte der Ereignisse zeigte einen Gegenstand, der sich verändert hatte (state change, SC), während die andere Hälfte einen unveränderten Gegenstand zeigte (no state change, NSC). In einer für die Versuchspersonen unerwarteten Gedächtnisaufgabe sahen die Versuchspersonen Bilder aus den Videos und sollten angeben, ob sie das Ende der Videos korrekt wiedergaben. Im Allgemeinen reagierten die Versuchspersonen am schnellsten auf Ereignisse, bei denen der Gegenstand unverändert blieb und die Videos vor dem Ende des Ereignisses gestoppt wurden, sowie auf Ereignisse, bei denen sich der Gegenstand veränderte und das Video nach dem Ende des Ereignisses gestoppt wurde. Darüber hinaus war in Experiment 2 die Erinnerung besser bei Videos, die abgeschlossen und im progressiven Aspekt präsentiert wurden ("*was chopping*"), und besser bei Videos, die im Verlauf waren und im perfektiven Aspekt präsentiert wurden ("*chopped*"). Dieser Effekt bezüglich des grammatikalischen Aspekts wurde nur beobachtet, wenn die Sprache nach dem Video präsentiert wurde (Experiment 2), aber nicht davor (Experiment 1). Die Auswirkungen von Sprache auf das Ereignisgedächtnis scheinen also davon abzuhängen, ob die Sprache vor oder nach einem Ereignis präsentiert wird.

Abschließend habe ich mich auf Bewegungsereignisse konzentriert. Frühere Studien zu Bewegungsereignissen haben gezeigt, dass Sprechende nicht-aspektueller Sprachen (potenziellen) Zielen mehr Aufmerksamkeit schenken als Sprechende aspektueller Sprachen. Ich untersuchte die Ereigniswahrnehmung bei Sprechenden einer aspektuellen (Englisch) und einer nicht-aspektuellen (Deutsch) Sprache, die auf einem Laufband über ein virtuelles Terrain liefen, das die Bewegung im wirklichen Leben simulierte. Bei dieser Aufgabe in der virtuellen Realität (VR) trafen sie sowohl auf Gegenstände entlang des Weges als auch auf potenzielle Endpunkte (oder Ziele). Durch den Einsatz von VR konnte ich die Wahrnehmung von Bewegungsereignissen in einer Umgebung untersuchen, die dem echten Leben nachempfunden ist. Die Augenbewegungen der Versuchspersonen wurde aufgezeichnet, um ein mögliches Bias gegenüber den Endpunkten festzustellen, da eine solche Neigung zuvor bei Sprechenden von nicht-aspektuellen Sprachen beobachtet wurde. Nach der VR-Aufgabe führten die Versuchspersonen eine Beschreibungsaufgabe und eine Beurteilungsaufgabe durch, beide mit Videoclips von Bewegungsereignissen. Bei keiner der Aufgaben wurden Unterschiede zwischen den beiden Sprachgruppen festgestellt. Der sprachliche Hintergrund führte nicht zu einem Endpunkt-Bias, so dass sich die zuvor festgestellten Forschungsergebnisse in einer Umgebung, die dem wirklichen Leben entspricht, nicht replizieren ließen.

Zusammenfassend zeigen die Ergebnisse dieser Doktorarbeit, dass Sprache das Verstehen von und das Erinnern an Ereignisse beeinflussen kann. Wir nutzen vor allem die Information, die Verben (z. B. "*to chop*") enthalten, um ein *event model* zu konstruieren, das Essenz des Ereignisses umfasst. Ein Verb definiert, wer was mit wem macht und dies hat Auswirkungen darauf, wie wir Ereignismerkmale wie Gegenstände verstehen und uns vorstellen.

Betrachtet man jedoch den grammatikalischen Aspekt, so sind die Ergebnisse gemischt. Obwohl Aspekt die Art und Weise beeinflussen kann, wie wir Ereignisse verstehen und uns an sie erinnern, treten solche Effekte nicht zwangsläufig auf.

Acknowledgments

Having spent more than seven(!) years at the MPI means having crossed paths with many, many lovely people.

Peter, I would like to thank you for taking a chance on me, for allowing me to take this PhD journey and make of it what I wanted. I filled my time with plenty of voluntary activities, decided to work on my PhD part-time, and you simply trusted that I'd get it all done anyway. And I did, just in time. Thank you for your feedback over the years and for your support.

Monique, we clicked straight away and I'm glad I made it past all the hurdles to become (one of) your first PhD student(s). It's been a bit of a wild ride, and I'm glad we shared the last years together. From discussions about grammatical aspect and aspects of my life, to the s*@t list to conference fun in Zurich and Boulder – you had my back, thank you so much!

David, thank you for your calm and reassuring nature throughout our VR project and its year-long revision phase. You made the start of my PhD easy and fun. Shame we never made it to that spinning class…!

Thank you to my reading committee Caroline Rowland, Christiane von Stutterheim and Carol Madden-Lombardi for taking the time to read this thesis and provide feedback. What an honour!

Kevin, without you the IMPRS just wouldn't be the same. You truly are there for everyone! We planned events, had big ideas and great conversations over the years. Your door was always open, thank you for all the motivating words.

Thank you, NBL group, for being such a supportive and fun bunch to work with and making my time at MPI great. Thank you, Ashley, Ellie, Yingying, Markus, Lydia, Evelien, Eva, René, Kris, Marvin, Natalia, Branka,… and all others who I've had lovely encounters with over the years. A special thanks to Birgit and Maarten, and Carolin, Ina and Tildie for heaps of practical support and always thinking along. Many thanks especially to all the PhD students in the group, who provided an open-minded environment to lament the failures and celebrate the success that make up PhD life. Our PhD meetings were such a blessing –thank you, Dan, for re-instating them (and for being my friend). Rowan, your office was one of my favourite spots to stop by for a lengthy chat about everything and nothing. Dick and Ksenija, you offered a hiding place and open ear when I needed it. Laura, thanks for your peaceful spirit, and Cas, thank you for wise input and a fun summer school. Teun, thank you for being so genuinely nice and enthusiastic. Alessio, thanks for bringing some good energy to our group. Anna, you came as a visitor and left as a friend. Thank you for creating projects together, it was great sharing some of the PhD journey with you. A special thanks goes to Iris for shaping my NBL days like no other: Iris, bedankt voor de leuke jaren(!) die wij op 333 hadden. Zonder jou had ik zeker nooit gedurfd Nederlands te praten en nu doe ik gewoon (bijna) elke dag. Zo fijn de PhD 'samen' met je gedaan te hebben en dagelijks over onze meisjes te kletsen.

Thank you, TG, for putting up with my technical mishaps. I am especially grateful to those (more or less) involved in my projects: Johan, Albert, Jeroen D, Thijs, and Reiner. Thank you also, Peter N and Ad, for welcoming me every now and then, and for sharing cups of tea. Thanks, Jeroen G, who was always up for last minute video edits (fun!), archive questions (the best!) and baby chats.

Thank you, Operations team, for helping me (and others) on so many levels. Over the years, I had the pleasure of working and chatting with all of you. Thanks to Rober, Jan and Pascal, who were always there when I needed my office re-arranged or found a lost bird in the corridor. Angela, Anique and Laura – when I needed my maternity leave arranged and contract changed, you made sure this important time in my life was taken care of. Thank you especially for thinking along with and supporting the PhD reps to improve working conditions for all PhDs at the institute. Pension? Pay rise? None of this would have been possible without you.

A large part of my PhD years was spent on voluntary activities I thoroughly enjoyed doing: Thank you, Federica and Eirini, for planning a successful career event together. Thank you, PhD reps, for being amazing team players: Miguel B, Limor, Dilay, Merel M, P and W, Fenja, and Julia – together we achieved some things that make me so happy and proud. To the current rep team: thank you for continuing the good work! Thank you, PhDnet Steering Group 2018: Jana, Severin, Greta, Angel, Maria, Ezgi, Mohamed and Ilka, I am so happy to have worked with you and am grateful we are still in touch. Karin, danke dir für das schöne letzte Jahr meines PhDs. Es hat wahnsinnig Spaß gemacht, zusammen mit dir am Gleichstellungsplan zu arbeiten und Ideen zu sammeln. Du bist eine Perle am Institut, wie schön, dich zu kennen.

My time at the MPI started as a student assistant. Little did I know that those slightly strange people I shared the Mini-Planck with in 2014, would become such important people in my life over the years. From TLA walks to going down and following the signs: Thank you, Alex, Elisa, Dieter, Twan, André, Paul, Peter W, Olha and Nanjo. You truly are the best –see you at lunch!

I had the pleasure of bumping into so many great people in the MPI corridors; often just a quick hello, sometimes a litte chat, other times a lengthy discussion. Thank you to my friend and calm soul Saoradh, Theresa, whom I shared language and gender thoughts with (and got to write about them, too!), Phillip, who always had an open door and honest advice, Patricia for coffee and pep talks, my BHV partner Francie, Miguel S for enthusiastic research ideas, Greta and Sara for their joyfulness, Maud and Thea for canteen chats, Laurel for being a passionate teacher, the BHV'ers, Tineke and Asifa, my first mentors, Edith, Joe, Paul, Arushi, Aslı, Gunter, Nina, Marlou, Erce, Katherine, Anne, Steve, Han, Judith, Annelies, Caitlin, Amie, Midas, Christina, Ine, the Equality and Diversity working group, Marjolein and the TalkLing team, my participants, and many, many more.

Danke an alle lieben Nicht-MPI-Menschen, die mich die letzten (wenn nicht schon all meine) Jahre begleitet haben: meine Eltern und meine Oma, Tanja, Christoph, Nikolaus und Konsorten, Annabelle, Robb, Ulla und Jochen, Carten, und Matze. Thank you to all old and new friends, close and afar, who shaped the last years (some doubling up as participants!): Pascal, Saya and the team, Caprice, Jo, Brendan, Ari, Hannah, Sofaya, Jackie, Gen, Poppy, onze buren Anne en Arnout, onze camping-buren Joske en Jordi, Aurelija, Annie, and more.

Thank you to my paranymphs: Ksenija, what would I have done without you those past years? We have gone through a lot this PhD and I am so glad you were there, your office my little haven. Your support means the world, and I always trust you'll be there for me. I'm so happy you are my paranymph – we did it! Sophie, you are a special someone. First my housemate, then my maid of honour and now my paranymph. I miss you here every day (this whole defence is just a ploy to see you again).

My deepest gratitude goes out to my loves. Three extraordinary people –my people!– two of whom were not around when I started this journey. Thank you for showing me every day what *really* matters in life. ♥

Curriculum Vitae

Julia Misersky was born on 21 December 1987 in East Berlin, Germany. In 2007, she moved to the UK, where she started a Bachelor's degree at the University of Sussex in 2009. After obtaining her BSc in Psychology (first class honours) in 2012, Julia worked as a research assistant at the University of Fribourg (Switzerland) and Macquarie University (Australia), before moving to the Netherlands in 2014. She completed a Master's degree in Cognitive Neuroscience with a specialisation in Language and Communication (cum laude) at Radboud University Nijmegen in 2016. Thereafter, Julia joined the International Max Planck Research School at the Max Planck Institute for Psycholinguistics as a doctoral researcher. Under the supervision of Monique Flecken and Peter Hagoort, she investigated the effects of grammatical aspect on our understanding of everyday events. Throughout her PhD, Julia was involved in several voluntary activities: As member of the 2018 Steering Group of the Max Planck PhDnet, Julia was elected to represent the 4000+ doctoral researchers of the Max Planck Society. In her home institute she worked in a team of PhD representatives actively improving the working conditions of her peers. In the final stages of her PhD, she was elected deputy of the equal opportunities officer at the MPI and co-founded the MPI's diversity working group. Since completing her thesis in 2021, Julia has taken up the role of Policy and Communications Officer at CLARIN ERIC, a European research infrastructure offering access to data, tools and services to support research based on language resources. She lives with her partner and two children in Nijmegen.

Publications

Misersky, J., Peeters, D., & Flecken, M. (under review). The potential of immersive virtual reality for the study of event perception. *Frontiers in Virtual Reality.*

Misersky, J., Slivac, K., Hagoort, P., & Flecken, M. (2021). The state of the onion: Grammatical aspect modulates object representation during event comprehension. *Cognition, 214*, 104744.

Misersky, J. & Redl, T. (2020). A psycholinguistic view on stereotypical and grammatical gender: The effects and remedies. In C. D. J. Bulten, C. F. Perquin-Deelen, M. H. Sinninghe Damsté, & K. J. Bakker (Eds.), *Diversiteit. Een multidisciplinaire terreinverkenning*. Deventer: Wolters Kluwer.

Misersky, J., Majid, A., & Snijders, T.M. (2018). Grammatical gender in German influences how role-nouns are interpreted: Evidence from ERPs. *Discourse Processes*, *56*(8), 643-654.

Misersky, J., Gygax, P., Canal, P., Gabriel, U., Garnham, A., et al. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods, 46*(3), 841-871.