Johannes Gerwien, Ines Marberg, Kristian Nicolaisen

What Are Events?

Abstract This study reviews research on the conceptual structures and cognitive mechanisms that underlie the human ability to cope with the dynamicity of the world, a phenomenon often subsumed under the term event cognition. It identifies four distinct theoretical perspectives in the literature: the participant-based view, the boundary-based view, the object-states view, and the event-layer view. The basic ideas that constitute each perspective are outlined, and the methods and empirical data that are typically presented in support of each perspective are discussed. After an evaluation summarizing differences and similarities, the conclusion is that likely only a unification of the four approaches makes it possible to capture all aspects of the phenomenon, including those related to perception, conceptual representation and linguistic encoding, and thus be of value to researchers across different disciplines. The final section offers some ideas on what such a unified approach might look like.

Keywords Event cognition; thematic roles; event boundaries; object states; event layers; cross-linguistic differences

Introduction

When people communicate with one another using language, they assume that their communication partners will understand more or less the same things by the words that are being said. In everyday life, slight variations in word meanings often go unnoticed. But in scientific communication, variation in the use of a specific term may have immense consequences. In the worst case, misunderstanding, misconception, and theoretical incoherence may emerge, even more so if the same term is used across multiple disciplines. However, investigating how a specific term is used by different researchers, acknowledging and incorporating terminological variation can have a more positive outcome, as well: a more fine-grained and inclusive understanding of a phenomenon.

One term worth reflecting upon how it is used across different disciplines is “event”. It frequently appears in the works of researchers investigating how

Gerwien, J., Marberg, I., Nicolaisen, K. (2024). What Are Events?. In J. Gerwien, I. Marberg & K. Nicolaisen (Hrsg.), *Die kognitive Perspektive. Wie Menschen über die Welt sprechen. Festschrift zu Ehren von Christiane von Stutterheim* (S. 225–255). Heidelberg University Publishing. https://doi.org/10.17885/heiup.1292.c20255

humans are capable of interpreting the ever-changing environment that they live in. Thus, it can be found in theories from a variety of fields that form (part of) the cognitive sciences: events play a role in research on perception, language, and memory. Although events may be treated as entities that happen among living organisms and inorganic materials in the world that surrounds us (Casati & Varzi, 2020), in research on cognition, events are taken as phenomena of the inner world. One of the issues in the cognitive sciences is how the mind generates and handles abstract representations of the outer world, whereas the ontological status of events is not necessarily the primary concern. The exact relationship between what is ‘within’ and what is ‘out there’ is apparently so complicated that thousands of years of philosophical thinking have not yet come to a definite, all-agreed-upon answer (Boden, 2006; Fuchs, 2017). For the purpose of this chapter, we will assume that an external stimulus can activate an internal representation and that an internal representation is critical for how the outer world is perceived and experienced at a given moment. So far, so good. But are event representations also evoked by internal simulation when one is contemplating experiences in the past, possible experiences in the future – or even imagining possible worlds while reading fiction? This paper aims to provide some cornerstones for the study of events in a way that reduces misunderstanding and thus provides a fertile ground for future research across disciplines.

If events are mental representations, how do they become available to real-time cognitive processing, what do they specify exactly, and how are they thought to play out in human cognition? These are the questions that we will address. Different theoretical perspectives can be identified in the literature, and each provides its own answers. Thus, we will start by introducing those views that we think can be grouped together and those that are unique enough to stand on their own. We will focus on the following four perspectives: the participant-based view, the boundary-based view, the object-states view, and the event-layer view. In the second part, we will summarize and evaluate the different views. We will try to point out some of the ideas that all approaches share, but also how they differ. In the last section, we will sketch in broad strokes how some of the differences may be overcome within a unifying framework. A disclaimer is in order, though: Due to space limitations, we can only present the core ideas of each of the four approaches, with only a limited level of detail. We hope that some of the references will help the interested reader to fill in the gaps that may arise.

Although it may be interesting for historical reasons, we do not think that it would be much of an advantage to tightly link every theoretical perspective we discuss to the specific discipline that it originated from because this might lead to the impression that the respective approach would be valid for that discipline only. Our goal here is to describe a broad spectrum of ideas revolving around the theoretical concept of event representations, and since all approaches, at their core, try to capture how the mind represents and processes the dynamic aspects of the world, they all reflect different aspects of the "same cognition" – at least this is the perspective that we chose for this chapter.

Four ways to approach event representations

The participant-based view

Participant-based views typically attribute their main idea to Charles Fillmore's (1968) *The case for case.* The theoretical core is evident in the following quote.

> The case notions comprise a set of universal, presumably innate, concepts which identify certain types of judgments human beings are capable of making about the events that are going on around them, judgments about such matters as who did it, who it happened to, and what got changed. (pp. 45–46)

Note that what Fillmore here calls "case notions" is referred to as "thematic roles" in subsequent work by most authors. According to Fillmore's conception, event representations comprise abstract atomic concepts. These concepts can be referred to by language and are thus represented separately from the domain of language. Each concept within an event representation corresponds to an entity that, in one way or another, participates in the represented event (hence the term "participant-based view"). Thus, two types of information are relevant in the participant-based view: the number of participating entities and how these entities are specified to relate to one another. For example, an event representation that may be referred to verbally by *The mother is waking up her girl* is a representation specifying two referents (*the mother*, *her girl*) as well as certain relations between them: one acts as an agent and a second one as a patient.

Dowty (1989) clarifies that Fillmore's original conception has not been perceived consistently in subsequent research. There are at least two versions of the participant-based view as derived from Fillmore's original theory. According to what Dowty labels the "thematic role type" conception, events are represented in terms of abstract role categories. If an event representation includes an agent, it refers to the same abstract agent concept as any other event representation that includes an agent. Similarly, if an event representation includes a patient, it refers to one and the same abstract patient concept as any other event representation that makes reference to a patient. Theories in this vein assume a finite set of abstract thematic roles from which every event representation can be constructed. Specific configurations of thematic role types result in different event types. For example, if an agent and a patient are part of the event representation, this may constitute a ‘causative event’; if there is an agent, recipient, and a theme, the event will be of the ‘transfer event’ type; and if there is a moving entity, a path, and possibly a landmark or a goal, then we are dealing with a ‘motion event’. Since the assumption is that there is a finite set of thematic role types, there is a finite set of event types. Note, however, that no agreement has been reached on which roles exactly constitute the assumed finite set, or how any particular role type may be defined, so that it is specific enough to be distinguished from other role types and general enough to cover a broader range of cases (Dowty, 1991; Rissman & Majid, 2019).

Things are different in what Dowty (1989) calls the “individual thematic role” conception, according to which each event representation has its own configuration of thematic roles. In eating events, for example, one referent is assigned the ‘eater role’ and another referent is assigned the ‘what is eaten role’. In borrowing events, one referent is assigned the ‘borrower role’, one is assigned the ‘borrowee role’, and one is assigned the ‘what-is-borrowed’ role. According to this theoretical conception, a high degree of similarity in participant configuration is assumed to hold primarily for exemplars of the same event type, e.g., for all eating events or all borrowing events. At the same time, the possibility that abstract role types exist is not completely ruled out, but it is not the core assumption.

Even though the core idea of event representations being configurations of thematic roles originated in linguistics, it has been and still is very prominent in experimental psychology today, and the way in which respective empirical studies are designed reflect the differences between the two interpretations of Fillmore’s original idea.

Studies that are motivated by the “thematic role type” conception typically aim to answer questions regarding the psychological reality of role types. This is done by attempting to obtain empirical evidence for cognitive biases associated with the processing of abstract roles across different events (Rissman & Majid, 2019). One line of research, for example, employs eye tracking during verbal description tasks as well as non-verbal decision tasks using visual stimuli. These studies focus on how people visually process pictures showing objects that can be interpreted as event participants. Typical research questions in such studies are: Do people show starting point preferences when looking at the experimental stimuli? How fast can people detect a given role? What factors modulate participants’ performance (Dobel et al., 2007; Glanemann et al., 2016; Griffin & Bock, 2000; Hafri et al., 2013; F. Wilson et al., 2011). In general, the results from such studies show that people can very quickly detect or identify an object that carries a specific event role in a depicted scene, that is, within a split second. This suggests that the interpretation of a visual scene as depicting an event is driven by the assignment of specific event roles to individual objects.

While thematic role assignment is apparently fast and automatic, not all roles seem to be processed equally: Some authors propose an agent-bias during processing, as agents have a special status in driving predictions about an event (Cohn & Paczynski, 2019; Kuperberg, 2021). Others propose that there is not only an agent bias but a cognitive hierarchy of event roles, placing agents before patients which are followed by goals and instruments. The hypothesis is that this hierarchy becomes apparent not only in frequency of mentions in free description tasks in which speakers can choose what to mention but also in preferences for the order of fixation, the time people need to move their eyes to the respective participant if the task demands it (F. Wilson et al., 2011) as well as in memory encoding and retrieval for different roles in the postulated hierarchy. So far, there is some limited evidence for such an event role hierarchy, although it is not clear whether this hierarchy is stable or situational. Recent studies suggest that it is malleable to a certain degree and dependent on various contextual factors such as priming (Sauppe & Flecken, 2021).

Although experimental research has a strong bias towards English and a few other languages, the cross-linguistic studies available may provide a promising road to investigate how thematic roles cluster with respect to their linguistic expressions in different languages and to draw conclusions about underlying (universal) conceptual structures, i.e., event representations (Flecken et al., 2015; Gerwien & Flecken, 2016; Rissman & Majid, 2019; von Stutterheim et al., 2020). However, an important question still under debate is whether differences that may be linked to the analysis of linguistic structures across languages indeed reflect the way people represent events at a domain-general level, that is, independently from the verbal domain (Gennari et al., 2002; Gerwien & von Stutterheim, 2018; Ünal et al., 2021).

To investigate the relation between linguistic and non-linguistic event cognition, some studies focus on individuals who are not assumed to exhibit biases stemming from linguistic experience, like apes, infants or congenitally deaf persons who were not taught any conventional sign language (Goldin-Meadow, 2003). Findings from this line of research suggest that the existence of different participant roles in events is deep-rooted in cognition and that agent-based event cognition even predates human cognition (V. A. D. Wilson et al., 2022; Zuberbühler & Bickel, 2022).

Studies in line with the tradition of the "individual thematic roles" conception follow a research agenda that focuses on the specific content of an event representation rather than the abstract relations between event participants. If we assume individual thematic roles to refer to semantic features that can constrain or bias referent selection or highlight or induce certain properties of the referent, then not every possible referent is a good candidate for a certain role, e.g., in an arresting event, a crook is less typically associated with the 'arrester-role' than a police officer. The concept of a typical arrester might even be automatically activated upon hearing the verb 'to arrest'. Ferretti et al. (2001), for example, argue that events are represented as schemas that specify prototypical referent concepts. In a series of masked priming experiments, they showed that specific verbs can prime prototypical agent, patient, patient-feature and instrument concepts. Vice-versa, typical referents for different semantic roles, including locations, also prime typical verbs (McRae et al., 2005). Thus, event schemata in this perspective are not only abstract structural relations between participants, but also clusters of object and action features.

It is fair to say that participant-based views dominate experimental research on event representation and cognition today. Despite the theoretical controversies surrounding the definition of a finite set of event roles, experimental psychologists with different backgrounds seem to favor this approach. One reason might be that it is relatively easy to develop an experimental design. Another reason might be that the event roles conception offers a simple vocabulary to describe what parts constitute an event representation. And once a researcher has words for the parts of a whole, it becomes possible to study the processes that create the whole from its parts.

The participant-based view has gained influence in the computer science community as well, where it serves as the theoretical backbone to implement mechanisms that enable machines to achieve natural language processing tasks. To some extent in the spirit of the individual roles conception, projects like FrameNet (Fillmore & Baker, 2009; Johnson et al., 2016) work on establishing a database of abstract 'frames'. These frames specify event types based on the relations that hold between referents that are interpreted to participate in the event.

The boundary-based view

The boundary-based view departs from the perspective of a third-person observer, and it holds that events are established by unitizing the continuous perceptual flow of information into discrete event units. In this view, events are time intervals that are delimited by event boundaries. Whatever lies between two boundaries is represented as an event unit. The following quote from Zacks et al. (2007) illustrates the idea:

> Thus, the system alternates between long periods of stability and brief periods of change. Periods of stability are perceived by observers as events and periods of change are perceived as the boundaries between events. (p. 275)

Event Segmentation Theory (“EST”, Zacks et al., 2007) is the most prominent representative arguing for the boundary-based view. Importantly, EST explicitly provides a mechanism by which humans create event representations; it thus explains how event representations emerge in real-time, that is, as resulting from automatic cognitive processing and associated neural activity. In EST, an event representation is called a ‘working model’, a term that highlights the fleeting, or transient nature of this specific type of representation. In short, a working model is a combination of abstract schema knowledge that is retrieved from long-term memory and the current input which the sensory system provides. A working model is kept activated in working memory, and once it is set up, a specific mechanism allows it to remain stable over some time. This is due to the assumption that new sensory input is predicted based on the working model and slight discrepancies between the model and the input are accommodated to some extent. However, if prediction error reaches a certain threshold, meaning if new input cannot be predicted well enough any longer, the current working model is abandoned and a new working model is established, again by retrieving a matching event schema from long-term memory. As studies suggest, this typically leads to the conscious perception of a breakpoint: one event has come to its conclusion and a new event begins (Zacks et al., 2001).

Experimental research grounded in the boundary-based perspective places great focus on the detection of event boundaries, namely by employing a task commonly referred to as ‘event segmentation task’ or ‘event unitization task’. Event segmentation is typically studied by presenting people with videos and asking them to press a button whenever one ‘situation’ ends and a new one begins (Newtson, 1973; Zacks, 2020). An analysis of when participants press the button while watching the videos reveals a relatively high agreement in where breakpoints are reported, although, of course, some variance between subjects arises. In same-group comparisons, this variance is attributed to the level of granularity that an individual chooses for detecting breakpoints. When choosing a fine-grained level for segmentation, subjects may detect many breakpoints, while choosing a coarse-grained level, fewer boundaries are reported in the same stimulus. A fine-grained or a coarse-grained segmentation can be induced by providing specific instructions. Importantly, breakpoints detected when choosing the coarse level coincide with some of the breakpoints that are reported when choosing a fine-grained level, highlighting that smaller event units are part of larger units. Such findings illustrate an important notion in event cognition research, namely that events can be seen as hierarchically structured. Macro-events comprise micro-events. Or the other way around: micro-events can be combined to form a macro-event (Bohnemeyer et al., 2007; von Stutterheim et al., 2020; Zacks & Tversky, 2001).

When correlating data obtained in the behavioral button-press task with neuropsychological data, as obtained with fMRI and EEG, two important observations can be made: 1) there are brain activity correlates of event segmentation; and 2) the same areas are activated both in active event segmentation and in passive viewing, where subjects simply watch a movie without being instructed to perform an event segmentation task (Zacks et al., 2001). These findings suggest that event segmentation happens automatically during perception.

A large body of research has explored how event boundary processing relates to information processing as well as memory encoding and retrieval in different groups of participants (for a recent overview, see Zacks, 2020). In these studies, participants typically watch video clips and their recognition and memory for what was shown in the videos is tested under different conditions. As event boundaries trigger the updating of a working model, information that occurs at a boundary receives special processing. For example, people notice changes in actors' clothing better at event boundaries than in intervals between boundaries (Baker & Levin, 2015). Similarly, objects that occur at boundaries are recognized better than objects that occur between two boundaries (Swallow et al., 2009). Integration or binding of information among elements occurs primarily within a segment between two boundaries, so cueing one within-segment element benefits the retrieval of other pieces of information whereas an intervening boundary may prevent these benefits (Ezzyat & Davachi, 2011). Across different segments, memory favors information that helps to establish causal links (Radvansky, 2012). In general, segmenting helps people to better recall fine-grained information about a series of events (Gold et al., 2017) and reduces temporal compression in memory (Jeunehomme & D'Argembeau, 2020). Event segmentation has also been studied in young vs. older participants and in healthy vs. Alzheimer's patients (Zacks et al., 2006).

To what extent language plays a role in event segmentation is not well understood to date. There are several ways in which the structure and the lexical repertoire as well as conventions of language use (pragmatics) may show effects. On the one hand, the way in which event schemas form over time during cognitive development could be influenced by the language one speaks, because how the members of a given language community use their language to talk about events may impact what information is most frequently clustered and what information is kept separate while talking about events. This could be labeled a long-term effect of language. On the other hand, a short-term effect of language could be attested if people formed event units differently when comparing how they segment input while talking about it in contrast to when they do not (Gerwien & von Stutterheim, 2022). That language has an impact on unit formation is suggested by the findings of a study by Gerwien & von Stutterheim (2018). They compared speakers of French and German in a verbal and non-verbal segmentation task and found that speakers of French were more likely to indicate event boundaries by button pressing when watching video clips showing a moving entity that changed orientation or direction in the course of the clips. Critically, French speakers also produced more assertions when describing the events spontaneously. The rationale of the study was based on the observation that speakers of so-called "verb-framed" languages, like French, must use individual verbs to refer to multiple path segments that a figure in motion traverses, while speakers of a "satellite-framed" language, like German, can combine multiple path segments in one assertion. However, the possibility that participants in the non-verbal task used inner speech to comply with the task could not be completely excluded, meaning that structuring the information for providing a button-press response could have been guided by internally creating linguistic structures. If that were the case, however, then the same argument could be made for any other event segmentation study.

To conclude our outline of the boundary-based approach, it should be acknowledged that compared to the participant-based view, which primarily focuses on the content of an event representation, the boundary-based view is much more concerned with the mechanism that creates event representations. Content-wise events are considered as having an internal structure, insofar as goals of agents and the outcome of actions play an important role (Kuperberg, 2021).

The object-states view

According to the object-states view, events are not taken as stretches of time per se, as in the boundary-based view, but rather as stretches of multiple times – or rather: 'time intervals' – for the individual object concepts that comprise an event. Thus, like in the participant-based view, events are composed of sub-components. There are two theories that may be subsumed under the term "object-states view", the Argument-Time Structure ("ATS") theory proposed by Klein (Klein, 1999, 2010)[1] and the Intersecting Object Histories ("IOH") theory, first fully spelled out in Altmann and Ekves (2019), but formulated in part already in Hindy et al. (2012). According to both theories, event representations specify how properties of objects develop over time. In the simplest case, an event representation specifies properties that are associated with only one object during an initial or 'source' time interval and during a resultant or 'target' time interval. The major difference between

[1] Klein's Argument-Time Structure conception was not intended to be a theory of event representation, but a theory capturing verb meanings and grammatical operators.

ATS and IOH is that IOH assumes that all time intervals between the initial and resultant state are part of the event representation (hence, “object histories”), whereas ATS does not make that assumption. Of course, event representations can be rich and specify states for more than one object. To illustrate the main idea from the perspective of ATS: In any event that represents some sort of transfer of possession, several objects co-occur in space and time. For example, in an event that may be referred to as *Mary threw John the ball*, there would be object concepts for *Mary*, *John*, and *the ball*. In addition, this particular event representation would contain an initial state for *Mary* which specifies the possession of *the ball* and a subsequent target stage at which the goal of the transfer – *John* – has come into possession of *the ball*. The specification of the properties that define each object state and how these states are associated via temporal and causal relations constitutes the minimal content of an event.

Since the argument-time structure of an event is a reductionist analysis of the descriptive and structural content, it is not a part of the ATS framework to account for transitional gaps between initial and final object states (e.g., a trajectory of a ball traveling through space and time before reaching a goal). IOH, on the other hand, posits a fine-grained view on object state representations by including all time intervals between the initial and resultant stage, thereby including transitional states. Both theories recognize that events are ensembles of objects (understood as entities with clear spatio-temporal boundaries, including living organisms) that undergo some form of state change over time. However, IOH explicitly states that our understanding of events implies the simultaneous activation of all object states associated with a specific event, i.e., the object histories in their entirety. In perception, this means that all object token states from the perceptual input must be bound into a coherent object representation and mapped onto semantic memory. From this view, generalizations such as participant roles are not the primitive components of event representation. Participant-roles result from patterns derived on the basis of perceived object changes.

The assumption of multiple state activation is backed up by both a priori arguments as well as experimental evidence. From a theoretical point of view, the activation of an object *history*, as opposed to the activation of a single perceptually salient or linguistically highlighted object *state*, is necessary in order to understand that a change has occurred. For instance, the content of the sentence *The burglar opened the door* is only recognized as an event as long as a previous state of *the door* (‘*not open*’) is part of the representation. The reasoning behind multiple state activation can be found in the framework of ATS as well. Here, object state activation is described as logical dependencies (so-called ‘H-connections’, named after David Hume), which imply that one object state would not be able to constitute a target state unless a previous initial state is counterfactually implied. A simple object state described by a sentence such as *The door is open*, in which *the door* has a single attribute (‘*open*’), is not recognized as an event since it does not imply a counterfactual relation to a previous state.

There is experimental evidence to support the idea of multiple object state activation. This concerns the following prediction: since events are understood as ensembles of object histories – meaning that representing an object implies representing multiple states of the same object at the same time – cortical activation patterns representing different object states must compete when a contextually appropriate object state is retrieved. In a study using fMRI, Hindy et al. (2012) found activation patterns suggesting such competition effects of different object states being simultaneously activated. In this study, participants read sentence pairs describing two different events containing the same physical object with the task to decide if the two sentences formed a coherent mini discourse or not. The first sentence described an affected object either as minimally changed (*The squirrel will sniff the acorn*) or as substantially changed (*The squirrel will crack the acorn*). The second sentence either described a preceding or proceeding interaction with the object (*But first, it will lick the acorn* or *And then, it will lick the acorn*). In order to understand the second sentence, a decision between retrieving the initial or target state of the critical object (acorn) had to be made. In this example, the acorn must be retrieved either in a sniffed or unsniffed (minimal change) or in a cracked or uncracked state (substantial change). Whereas sniffing an acorn would leave the substance of the acorn intact, cracking an acorn would change it. Since the comprehension of an interaction between the squirrel and the acorn would imply deciding which object state of the acorn is contextually relevant, a subsequent interaction involving a substantial change in state of the acorn would induce a greater semantic conflict than a minimal change. This means that, given our world knowledge of the affordances of an acorn, the semantic conflict would then arise in the situation in which the squirrel would lick a cracked acorn. The results of the study showed activation patterns in the substantial change-condition indicating competition in the retrieval of different object states. Further experiments confirmed that the effect is not due to the processing of specific lexical elements, but indeed to the change of a specific object in the context of the event (Experiment 2 in Hindy et al., 2012; Solomon et al., 2015). In addition, results from two further studies suggest that the competition effect is indeed linked to the subsequent reference to the object (Kang et al., 2020a; Prystauka, 2018). Reaction time studies by Kang and colleagues (2020b) and Horchak and Garrido (2021) using a picture-sentence matching task confirmed that after a sentence implying a change in object state, initial and resulting state remain activated regardless of the object being mentioned again in a second sentence. The underlying neural mechanisms were investigated in more detail by Hindy et al. (2015) in an fMRI study. They concluded that different object states are represented as being different in the primary visual cortex, but as the same stable object in the left ventral posterior parietal cortex, and that the mechanism observed in the study by Hindy et al. (2012) supports the selection of the sensorimotor representation relevant in the current context.

Another line of research is less concerned with the degree of change that an object undergoes, but rather with the complexity that arises per se if a verb denotes multiple object states. Gennari & Poeppel (2003) compared eventive (multiple object states) and stative (single object state) verbs using reading times and lexical decision reaction times and found that eventive verbs take longer to process, which they explained by the more complex event structure. Gerwien (2011) followed a similar logic and compared reading time for intransitive verbs that do or do not denote a change of state. Results show longer processing times for more complex verb meanings.

In summary, the object-states view defines events with focus on their unfolding over time. Unlike the boundary-based view which is concerned with the cognitive mechanisms segmenting events into units, the object-states view sets out to explain how the representational content of these units becomes available over time. Representational content can be described as bundles of object states co-occurring in spatio-temporal proximity. This view on events can be described by principles laid out by ATS and IOH. While ATS models the minimal descriptive event content (i.e., object states) that can be referred to by verbs as well as the logical and temporal relations between object states, IOH focuses on perceptive mechanisms of binding multiple object states into coherent representations. The latter theory then shifts focus from generalized static knowledge onto the neurocognitive processes involved in abstracting generalized representations from primitive building blocks during event conceptualization.

The event-layer view

The event-layer view starts with the observation that things in the world frequently happen in parallel. Bennett (2002) refers to this observation as events being ‘thick’. To illustrate this view, one can imagine that when a leaf falls from a tree, the leaf would be moving downwards while possibly also rotating around itself. An object can be in motion on some path, for example, moving downwards, while at the same time exhibiting a specific manner of moving, for example, rotating. Importantly, the relation between falling and rotating in this example is not hierarchical in nature. Falling is not a subevent of rotating and rotating is not a subevent of falling. While events can be and often are hierarchically embedded in a taxonomy of causal links (see Goldman, 1970; Löbner, 2021 and "boundary-based view" section), the event-layer view captures a different kind of complexity in event representation, i.e., parallel occurrence with no explicit causal relation.[2] The relevance of parallel occurrence for questions of event representation is, for example, laid out in the work of von Stutterheim and colleagues (Gerwien & von Stutterheim, 2022; Gerwien & von Stutterheim, 2022; Lambert et al., 2022; von Stutterheim et al., 2020; von Stutterheim & Gerwien, 2023). The main idea behind what this research group studies under the term 'event layers', is that an observer of a scene selects one specific dimension – one 'event layer' – as a starting point for the construal of an event representation. The choice in turn has consequences for unit formation. Imagine a person running toward a train station and then entering it. Choosing the 'path layer' for event construal would lead to two event units in this case, one that represents the phase of approaching and one that represents the phase of entering the train station. Choosing the manner layer, on the other hand, would yield only one event unit which would represent the phase during which the figure continuously exhibits the same manner of motion without changes (e.g., running). Von Stutterheim and colleagues explore the event-layer view in several studies, in which they collect spontaneous verbal descriptions from speakers of different languages who respond to short video clips of real-world situations. The analyses of the verbal responses reveal patterns of what information speakers select for expression, what information they omit, and over how many clause-sized linguistic units (assertions) they distribute the information they provide. The authors interpret these patterns in verbal descriptions as to reflect preverbal event representation, that is, the representation of information that feeds into linguistic encoding.

Note that the event-layer view may not be restricted to the domain of motion events. Viewing the construal of an event representation as selecting one privileged layer as a starting point may apply to causative and other event types as well. Take a sentence like *Someone is folding a paper airplane*. Using this sentence in English refers to a manner in which an actor handles a piece of paper (folding), and at the same time it refers to the changing of a sheet of paper from one shape into another, which can be described as the intentional goal of the actor. While there is nothing special about expressing information

[2] The existence of non-causal relations between simultaneously occurring actions has been identified by Goldman (1970) ("simple generation"; p. 31), exemplified by the parallelism of the acts of *hitting the tallest man in the room* and *hitting the wealthiest man in the room*, by which the same person is being referred to. In line with Goldman, Löbner (2021) explains such relations as a "constellation of facts" (p. 270). However, there is no further elaboration on how non-causal relations between parallel events occupying the same spatiotemporal zone should be modeled in abstraction from human action.

on the intentions of the actor and the manner by which the intentional goal is being reached for speakers of English or German, many languages do not allow speakers to refer to both layers – the manner layer, and the intentional layer – at the same time, i.e., in a single assertion. For example, to express the same information conveyed by the example above, a speaker of French would have to use the verb *plier* (to fold) and the verb *faire* (to make). If she intends to produce only one assertion, either the folding (manner) or the paper plane-making (intentional goal) needs to be omitted.

The event-layer view, as far as it has been made explicit in the published work of von Stutterheim and colleagues up to this point, should not be regarded primarily as an attempt to identify a finite set of layers that may comprise 'thick' events. So far, the types of layers that may be identified include the path (in motion events), the manner, the intentional, and a 'witness' layer (von Stutterheim et al., 2020), although the authors emphasize that this list is not exhaustive. Also, linking specific information selection patterns that may be found in different languages with specific features of the grammar and lexis in those languages is in our view not the most relevant aspect of the approach. What we think is worth underlining here is that events are regarded as representations that are *construed* by actively directing attention to information, either by the need to fulfill the current task requirements or by a default mechanism if there is no specific task. Thereby, some information receives a privileged status, whereas other information is defocused. Note that only after the choice for a specific layer or layers has been made, it can be determined which objects will be represented as participating in a given event and which objects will not. If we use the terminology of the participant-based view, selection of an event layer determines whether an object may be the agent or a moving entity, or whether an object will be a theme or an instrument, and so on. The choice for the layer determines what relations are relevant between which objects. Similarly, from the perspective of the object-states view, choosing one layer determines which object and which object states will ultimately be part of the event representation. Thus, assuming event layers, at least as an epistemological tool in the study of event representation, makes it possible to pre-structure the seemingly infinite possibilities of what information will be represented temporarily and what information will not be processed further.

Under a set of rather specific assumptions about the general architecture of the cognitive system and corresponding assumptions about the levels at which information are represented, i.e., a distinction between verbal and non-verbal levels of representation (Lupyan, 2012; Paivio, 2014; Van Dijk & Kintsch, 1983; Wolff & Holmes, 2011), the event-layer view may seem to imply that there is some additional, maybe even a 'richer', representation that provides the 'material' from which information is selected and represented specifically for verbal expression. If one makes such assumptions, the question may seem justified as to whether speakers with different languages do indeed represent different event units at a non-verbal level of representation, or whether they simply represent the same events, but refer to them in different ways. We will address the question of how many levels of representation researchers assume in some more detail below. As far as the event-layer view proposed by von Stutterheim and colleagues goes, there is only one level of representation, and that representation holds information as provided by the sensory systems and information from activated event schemas, which to some extent are modulated by language (see Gerwien & von Stutterheim 2018, and above). Support for this view comes from studies on bilinguals and highly advanced L2 speakers, which suggest that some of the patterns in information selection during event construal prevail when participants use a language other than their native language to verbally respond to visual stimuli. Thus, what drives ad hoc event construal for verbalization is not necessarily influenced by the grammar and vocabulary of the language currently in use but may rather be driven by deeply-entrenched patterns of attention allocation that guide the information selection process and that have formed as a consequence of native language use. Another piece of evidence comes from the already mentioned study by Gerwien & von Stutterheim (2018), where the main manipulation concerned whether a person or inanimate object did or did not change direction / orientation in the course of short video clips. Results showed that French speakers were more likely to produce multiple assertions than German speakers in a verbalization task, and that they were also more likely to indicate an event boundary by button-pressing in a non-verbal event segmentation task. Von Stutterheim and colleagues argue that speakers of different languages make use of what they call ‘attentional templates’ that guide information selection and consequently the verbal encoding process. In their view, these attentional templates are stored as event schemata in long-term memory. Which of the available templates is used as a default in a given situation may vary from language to language (Gerwien & von Stutterheim, 2022; Lambert et al., 2022).

Summary and evaluation

In the last sections, we provided an overview of four perspectives on event representation. In the participant-based view, events are defined by the number of referents participating in the represented event and the relations between them. In the boundary-based view, events are discrete units segmented out of the continuous perceptual stream. In the object-states view, events are seen as different states of objects associated with one another. And in the event-layer view, events comprise different qualitative dimensions linked to one or several entities. Given these apparently quite distinct ideas on event representation, the question arises: Do all approaches mean the same when using the term "event"?

All four approaches assume that, at a theoretical level, events have 'a right to exist' as conceptual representations, next to other representations such as object representations. In this, all four approaches take events to be cognitive units. Considering events as cognitive units implies that, at least at some point during real-time processing, they are available to the cognitive system as a whole. Similar to objects, which may be analyzed as consisting of subparts – a human body for example comprises a torso, a head, legs and arms –, but which, for the sake of being interpreted as a whole, appear to be integrated at a given moment in time, events also integrate components to form a whole. This has implications for questions related to the processing, memory, and verbal encoding of events, as well as to the acquisition of event processing abilities during cognitive development and the potential loss of these abilities in cognitive decline.

The four approaches, however, differ with respect to the 'size' of an event unit, that is, regarding how many components, and thus, how much information can be integrated to form one unit. On the one hand, the participant-based view, the event-layer view and Klein's Argument-Time-Structure (ATS) theory, as one representative of the object-states view, tend to restrict the amount of information that can be integrated to form an event to what can be expressed in one sentence. That is, while the duration of the temporal intervals that constitute events event can vary freely – potentially ranging from something that can happen in the external world in a split second to something that can take place over minutes, days, months or even millenia –, the subparts that are to be specified and integrated depend on, and are restricted in number by what can be expressed in a sentence that includes a specific verb. The main components are the referents that are directly associated with the verb as the verb's syntactic arguments. On the other hand, neither the boundary-based view nor Altmann and Ekves' Intersecting Object Histories (IOH) theory in principle assumes any specific unit size in terms of the number of subcomponents that may be integrated to form one event unit. In IOH, units are defined by binding mechanisms, including the binding of objects to event schemata. Thus, whatever components the binding mechanisms are applied to will be part of the event unit. In the boundary-based view, units are defined by perceptual boundaries at one of potentially many different (hierarchical) levels whereas there are no principled restrictions as to the number or type of subparts.

Importantly, each perspective acknowledges the need to be able to explain why people can recognize events as being in a certain way similar to events that they have encountered previously and why people use similar linguistic devices to talk about similar events. Thus, all approaches make a distinction between abstract event knowledge on the one hand and specific or instantiated event representations on the other hand. Obviously, it depends on the theoretical approach of what can be described as the content of an event representation to make explicit what ‘abstract event knowledge’ consists of exactly. Abstract knowledge may be taken as knowledge about abstract event roles or as knowledge about prototypical participants of types of events, including not only typical animate objects (agents and patients), but also typical locations and typical additional objects, such as instruments. Or abstract knowledge may be knowledge about how a person typically reaches a specific goal, e.g., knowledge about how an agent’s intentions play out. Or abstract event knowledge may be viewed as knowledge about how features of objects typically change over time.

Given that all approaches make a distinction between abstract event knowledge and specific instantiations of that knowledge, all four approaches, implicitly or explicitly, locate event representations in working memory. Note that we mean here the “generic definition” of working memory as provided in Nelson Cowan’s widely perceived review article “The many faces of working memory and short-term storage”, according to which the term refers “… to the ensemble of components of the mind that hold a limited amount of information temporarily in a heightened state of availability for use in ongoing information processing.” (Cowan, 2017). If it is acknowledged that events are units in working memory, then events are fleeting, or transient representations. Events are representations at a specific moment in time that vanish, unless kept activated by some cognitive mechanisms. However, the literature is not too explicit regarding whether the term ‘event’ should be restricted to this notion, or whether the term should be applied also to representations stored in episodic memory, i.e., to knowledge of specific personal experiences that have been represented as specific events at some point in the past, but that is not currently activated. If event representation is considered as a constructionist process, there seems to be no need for that, because recalling a personal experience could simply be thought of as (re-)creating a specific transient event representation from smaller pieces of information.

Although to different degrees and in different ways, the four approaches imply that the formation of an event representation in working memory requires a control mechanism by which relevant information is selected and shielded from interference by irrelevant information. Despite the fact that the term ‘attention’ is heavily debated in current research, with some authors even calling for its abandonment (Anderson, 2011; 2023), we will nevertheless use it here. In this, we view working memory as a storage component and attention as the processing that acts upon the temporarily stored information. It is hard to imagine how the construal of a cognitive unit from ‘smaller parts' by means of integration, and how keeping the resulting representation activated over some time could be achieved without assuming any type of an attentional control mechanism. To consider attention a necessary cognitive prerequisite to study event representation and event cognition is imperative to tackling a variety of issues that could not be resolved otherwise. First, attention prevents random shifts between different granularity levels during event unit formation (see section 'The boundary-based view'). Without attending to one particular level of granularity at a given moment in time, it would be impossible to interpret a dynamic scene coherently or to find words to talk about it. This is also evident from the fact that humans do not seem to automatically establish every relationship between a given event and its embedding in, or relation to, the global affairs of the world at all times (cf. the micro-/macro-event distinction). Second, attention is required to select relevant information in the sense that the layer-view suggests. For example, while observing traffic with the goal of crossing a street, one most likely will attend to the movement of vehicles (the path layer) and not that the drivers of the vehicles have intentions of moving their vehicles to certain destinations (the intentional layer). Similarly, the manner in which a particular action is performed may not be represented obligatorily as part of an event: If you hear someone say that person A woke up person B, you may represent the change of state associated with person B but not how that change of state was brought about, e.g., by hitting a metal pot with a wooden spoon, or by poking, whistling, or calling a name. Choosing a hierarchical or qualitative level is relevant for event construal both based on the immediate visual input as well as for event construal from memory. Thus, even though at all times multiple dimensions are available to extract information from in order to form an event unit, the selection depends on the goal or task at hand, for which on theoretical grounds the assumption of an attentional selection mechanism seems to be required. Yet another reason for why an attentional control mechanism must be assumed to make event representation possible can be subsumed under the term 'perspectivization'. If working memory, as a temporary storage contains the components that make up the event, then these components can be 'profiled' in different ways. In a specific event, the focus may be on the causer or on what is being caused, as illustrated by comparing the event representations referred to by *Tom was woken up by a loud noise* and *The loud noise woke up Tom*. A similar case is illustrated by a comparing what is being referred to by *The hunter is chasing the deer* and *The deer is fleeing from the hunter*. Since information selection, information integration and information manipulation are inherent to some extent in all of the presented approaches, some theoretical assumptions about temporary information availability (storage) and processing (attention) must be made.

Levels of representation

How many levels of representation must be assumed to understand event cognition as the fundamental human ability to structure information in order to make sense of it and to communicate about it, no matter whether the origin of the to-be-structured information is perception, episodic memory, imagination, or language? It must be attested that there is no consensus on whether the modality independence implied by the question is a valid assumption at all. There are at least two reasons: One, researchers have been approaching event representation from individual research fields, and therefore from different directions, e.g., from visual perception or from language. Two, some researchers study events in isolation and others in context. The first reason leads to two very different views: On the one hand, there is the assumption that event representations are modality-specific, that is, event representations construed for linguistic encoding and events resulting from (visual) perception are not identical (e.g., Papafragou, 2015; Papafragou et al., 2008). On the other hand, there is the assumption that event representations are not specific to one modality, that is, the view that events are represented at a domain-general level. The second reason concerns the relation between events and another representational device termed 'mental models'. We will start with addressing the latter.

While many researchers distinguish between events on the one hand and mental models (Johnson-Laird, 1983) or situation models (Van Dijk & Kintsch, 1983) on the other hand, a few researchers in the domain of event cognition assume a homology between situation/mental models and events. For example, Speer et al. (2007), in a study based on EST, write: "As events structure visual activity, situation models necessarily structure narrated activity". First, it is important to note that "... although there were many differences in the ways we [Van Dijk/Kintsch and Johnson-Laird] used the notion of a model [in the terms 'situation model' and 'mental model'], the fundamental idea was the same." (Van Dijk, 1995). Thus, what is said in the following regarding the more general theoretical concept of a 'mental model' applies to the somewhat more specific concept of a 'situation model', as well. A 'mental model' is traditionally seen as a representation that is a complex mental simulation of relevant aspects of the world at a given point in time, which integrates different pieces of information from multiple sources, including information provided by the perceptual apparatus and information from long-term memory (scripts and schemas). This means that mental models are modality-independent in the sense that they combine information from different modalities into one representation. Note that it is a separate question whether this implies a multimodal (modality is preserved) or amodal (modality is not preserved) format. Humans use mental models to tackle all sorts of higher-order cognitive tasks, such as text / discourse comprehension (Kintsch & van Dijk, 1978; Van Dijk, 1995; Van Dijk & Kintsch, 1983), reasoning and problem solving (Craik, 1943; Johnson-Laird, 1983). Mental models are updated whenever new information becomes available, and thus, a mental model is typically conceptualized to integrate multiple individual events as they become available as new perceptual input or from memory retrieval. This implies that, in theory, event representations, as they are understood by the majority of researchers in the domain, and mental models cannot be representations at the same level. There cannot be a homology. But how, then, do they relate to one another? On the one hand, events can be understood as some of the building blocks of mental models, though, mental models integrate other 'non-eventive' information such as visual properties of a scene like the color of the sky, or factual knowledge such as knowing that it is dark at night as well. In other words, event representations could be seen to constitute some of the 'material' that mental models are made of. On the other hand, a mental model may provide the information from which one or several events can be constructed, e.g., for verbalization. If you have experienced something extraordinary yesterday, then what happened to you will be represented as a whole in the form of a mental model. If you want to tell your friends about it, you may select relevant pieces of information from your mental model and create one or multiple event representations that allow you to refer to different parts of it. Whether events are the building blocks of a mental model (among further non-eventive information), or whether a mental model provides the information for event construal (among further non-eventive information) simply depends on the perspective: Perceiving information that can be interpreted as an event can create or amend the current model, whereas, if the cognitive task is to evaluate aspects of the current mental model, e.g., in decision making, or to communicate about what is in the model, an event representation may be constructed from the information in the model. From this perspective, events are cognitive devices that interact with mental models in the service of the current task.

We now return to whether event representations are modality-specific or modality-independent. The boundary-based view (EST) and IOH – one of the representatives of the object state view – both focus on the mechanisms at play that underlie event representation. In this, both approaches do not explicitly differentiate between any modality-specific levels of representation. As Altmann and Ekves (IOH) put it: "There is little difference between directly experiencing [an event] and learning of it through language; yes, there are differences in detail (and goals), but by 'little difference' we mean in respect of the mechanism by which the tokenized representations come about" (Altmann & Ekves, 2019). In contrast, the large majority of researchers adopting the participant-based view assumes that events are represented at multiple levels of processing and that it is possible to isolate different representations as the result of different processing steps, e.g., in the course of visually perceiving an event to the linguistic encoding of it (see section 'the participant-based view'). To illustrate this way of thinking once more, imagine a situation in which someone is presented with a video clip with either the task to describe it ('verbal task'), or to make some judgment about it or memorize it for later recall ('non-verbal task'). In each of these scenarios, visual information first needs to be taken in via the sensory system. Some researchers assume that the result of information uptake leads to a representation of the visual input and based on that representation, further task-dependent representations may need to be generated. In the case of the verbalization task, the task-dependent representation would correspond to a semantic structure serving as a compatible input of the linguistic encoding system (following Levelt [1989], a so-called 'message'). In the case of a non-verbal task, no such representation is assumed to be required. The comparison of eye movement patterns registered during such verbal and non-verbal tasks have been reported to show differences (Griffin & Bock, 2000; Papafragou et al., 2008). In addition, differences arise when eye movement patterns are compared in verbalization tasks between people with different native languages, whereas differences are absent when comparing groups of speakers with different languages in non-verbal tasks. Such findings are often interpreted as evidence for the 'multiple-levels'-view (Papafragou et al., 2008; Trueswell & Papafragou, 2010). However, many have pointed out the methodological pitfalls associated with the goal to obtain evidence for non-verbal event representation in such a way, as well as problems with the underlying theoretical conception. For one, since the assumption is that different levels of event representation exist independently of different tasks, and different tasks only serve the purpose to tap into these different levels representations, it is unclear how one can dissociate between task-effects and 'representation-effects', so to say. It does not seem unreasonable to assume that the mind can generate representations specifically for the tasks at hand, which does not necessarily imply different levels of representations. An in-depth discussion of all issues is not possible here due to space-limitations, but we refer the interested reader to, for example, Wolff & Holmes (2011), Lupyan (2012) and Gerwien, von Stutterheim & Rummel (2022).

If one appreciates the theoretical concept of mental models and how events can be conceptualized to relate to them, namely in the form of transient representations that can operate on the information in the mental model, then the question whether events are represented at multiple levels of processing loses its theoretical relevance. Event representations are essentially associations of different pieces of information, which may be available in a multi-modal representational format. However, a different question may be formulated: Is information that serves as input to the mental model already structured in the format of an event before its integration, or are events only formed within the mental model?

What may a unified theory look like?

All four views that we chose to present here highlight different aspects of event representation as a cognitive phenomenon. Thus, all four approaches potentially specify the ingredients for a unified theory that can provide both a universally applicable format of event representations as well as the processing mechanisms involved in event cognition.

One starting point towards a unified approach would be to translate those ideas that primarily focus on the content and structure of event representation into a common format. One way to do so would be to assume that, in terms of their contents, events are nothing more but representations in which attributes of objects are attended to over a period of time. In this, the term ‘attribute’ applies to all kinds of inherent (color, shape, animacy, etc.) and extraneous (conceptual salience, intention to reach a goal, etc.) features that can be linked to objects. This includes features that can be subject to sensory perception (e.g., motion, change of integrity) and features that can be assigned (e.g., volition, intention). For the sake of the argument, let us refer to the first as ‘attribute perception’ and to the latter as ‘attribute assignment’. In both cases, attributes define a domain for attention allocation. By attending to an attribute of an object over time, the values of that attribute may either change or stay constant. One may refer to this as ‘attribute value tracking’. An event representation must at least bind two values of the same attribute to one object. Note that parts of this idea were laid out in Miller & Johnson-Laird’s influential book *Language and Perception* (1976). The general view just outlined may provide the basis for everything that can and must be said about the content of an event representation, even for very complex ones. In addition, it offers several advantages, which we will describe in more detail below.

Second, it would be necessary to adopt and adapt the mechanism laid out in the Event Segmentation Theory, that describes how generalized event knowledge is activated to create a ‘working model’, to the idea of ‘attribute tracking’. This should not be too complicated as quality changes (changes of attribute values) are already taken as what determines the deactivation of a current and the activation of a new event schema in EST, i.e., what constitutes the perception of a boundary between two events. However, previous research conducted from the perspective of EST has been more concerned with how the mechanism affects the construction of situation / mental models rather than event representations in the sense we laid out here. This is where adaptation would be required.

Let us return to the idea of attribute assignment, attribute perception, and attribute tracking. First, allowing both attribute perception and attribute assignment as equal cognitive operations relevant in constructing event representations opens up the possibility of bottom-up and top-down driven modes of event cognition. In other words, it allows perceptual features to be the starting point to generate an event representation, and at the same time, it leaves room for conceptual guidance. This view allows interdisciplinary theories of event cognition to overcome the great (artificial) divide between perception and cognition. If attribute tracking is assumed to be at the heart of event cognition, it should not matter too much whether attention to attributes is *triggered* by perception or whether it is *assigned* driven by a task-goal. The resulting representation can have the same format.

Second, attending to object attributes over time (as opposed to viewing them as static features) is what theoretically differentiates object representations from event representations. Identifying the tracking of attribute values over time as the core of event cognition allows us to do away with the distinction between "events in the narrow sense", where there needs to be a qualitative change ('to wake up'), and "events in the broader sense" ('to sit at the table'), where an actual (perceivable) change is not the defining criterion. Allocating attention to an attribute value during a time interval without identifying any change would allow to create a dynamic representation, meaning an event, and not a representation of a state of an object.

Since in many cases event representations integrate the tracking of attributes of *multiple* objects over time (e.g., 'event participants'), an explanation is required of how this is achieved. One possibility is to assume pre-made structured representations to which attributes of multiple objects can be linked as elements of that structure – an idea which is in part laid out in the Argument-Time Structure theory (ATS, Klein, 1999). It may be possible to identify a finite set of time interval configurations to which objects and object attribute values can be bound, with the least complex one comprising only one interval for one object, and more complex ones comprising multiple intervals for multiple objects. It seems likely that the complexity of time interval configurations has an upper limit, which is imposed by the resource limitations of the cognitive system, most importantly the limit of items that can be maintained in working memory simultaneously. Assuming pre-made structured time interval configurations to which relevant object attribute values can be linked provides an easy solution to the problem of how the tracking of attributes of multiple objects results in an integrated representation. Figure 1 summarizes our proposal.

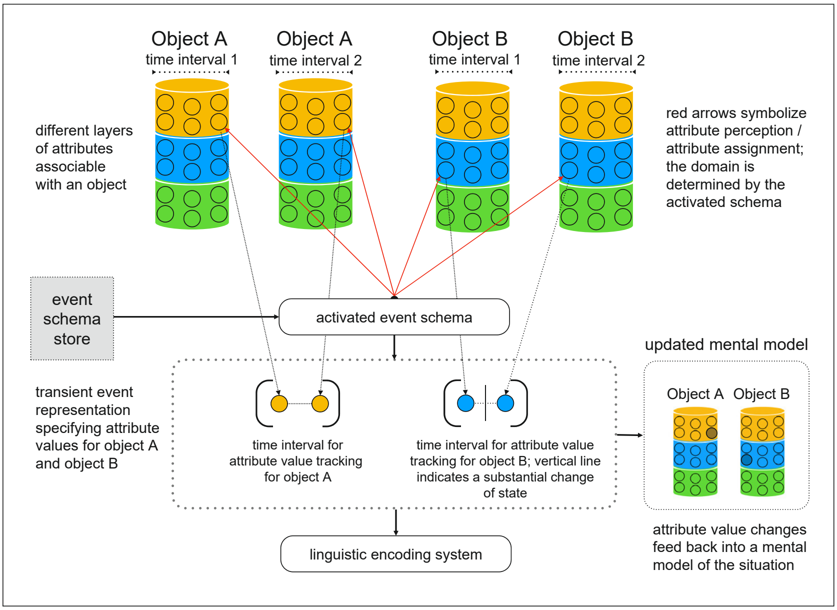
Object A
time interval 1
Object A
time interval 2
Object B
time interval 1
Object B
time interval 2
different layers of attributes associable with an object
red arrows symbolize attribute perception / attribute assignment; the domain is determined by the activated schema
event schema store
activated event schema
updated mental model
transient event representation specifying attribute values for object A and object B
time interval for attribute value tracking for object A
time interval for attribute value tracking for object B; vertical line indicates a substantial change of state
Object A
Object B
attribute value changes feed back into a mental model of the situation
linguistic encoding system

Figure 1. Illustration of a unified conception of event representation; the top part illustrates attribute perception, attribute assignment, and attribute value tracking, which all interact with event schema knowledge (red arrows); the area with the dotted contour illustrates an event representation; the right part illustrates how a specific event representation can update a mental model.

Conclusion

We can't do without events, or can we? What researchers attempt to capture with the theoretical concept behind the term 'event' is fundamental to human cognition: the ability to survive and live in an ever-changing environment. The investigation of how humans perceive, remember, and communicate about events, and how these skills are acquired, however, poses some serious challenges to the interdisciplinary research of the phenomenon. As the mind (and its implementation in the brain) has been studied by different disciplines for almost two centuries now, each discipline brings its own epistemological instruments to the table – methodological and theoretical – which inevitably leads to a situation where the same phenomenon is latently associated with different discipline-specific concepts. For example, one cannot study events in linguistics without acknowledging basic linguistic notions such as the fact that words are combined into clauses and sentences, which provides the linguistic units onto which events must be mapped. Similarly, if events are studied in cognitive psychology, the processes by which the mind generates event representations must be linked to the assumed general cognitive architecture, including core components such as the so-called "executive functions" (e.g., working memory / attention).

At some point during the preparation of this chapter and the discussions that went along with it, we suspected that the term 'event' suffers from what is known as 'conceptual fragmentation', as many other theoretical terms especially in cognitive science, like for instance the terms 'representation', '(working) memory', 'attention', and even 'cognition' itself, among others. Conceptual fragmentation refers to cases where "… (i) a certain term, originally widely assumed to enjoy a single meaning, has been found to have multiple distinct meanings no one of which is privileged, and (ii) different definitions are adopted for different theoretical uses." (Taylor & Vickers, 2017). However, if our suspicion was correct, we have hopes that different disciplines can find common ground regarding the concept of 'events'. We think that it is possible to develop converging ideas on the underlying theoretical concept and use the term in a coherent way. Our goal here was to pave the way for a better understanding between disciplines by trying to extract four main approaches to event representation, crystallize each one's core ideas, and so, for one, create awareness of the different facets of the phenomenon, and two, outline how it may be possible to combine them into a unified framework.

To conclude, event representation is a research topic that cuts through all aspects of cognition. Therefore, it is an arena where different disciplines can come together to learn from each other, present their unique perspectives, but also to solidify or challenge some of their specific notions and assumptions. Theoretical concepts that do not (fully) work in explaining event cognition may require re-evaluation. Finally, events are an interesting field for evaluating meta-questions in the field of cognitive science, such as representationalist and anti-representationalist views on cognition (and all positions in between).

References

Altmann, G. T. M., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*, *126*(6), 817–840. https://doi.org/10.1037/rev0000154

Anderson, B. (2011). There is no Such Thing as Attention. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00246

Anderson, B. (2023). Stop paying attention to "attention." *WIREs Cognitive Science*, *14*(1). https://doi.org/10.1002/wcs.1574

Baker, L. J., & Levin, D. T. (2015). The role of relational triggers in event perception. *Cognition*, *136*, 14–29. https://doi.org/10.1016/j.cognition.2014.11.030

Bennett, J. (2002). What events are. In R. M. Gale (Ed.), *The Blackwell Guide to Metaphysics* (pp. 43–65). Blackwell.

Boden, M. A. (2006). *Mind as machine: A history of cognitive science*. Oxford University Press.

Bohnemeyer, J., Enfield, N. J., Essegbey, J., Ibarretxe-Antuñano, I., Kita, S., Lüpke, F., & Ameka, F. K. (2007). Principles of event segmentation in language: The case of motion events. *Language*, *83*(3), 495–532. https://doi.org/10.1353/lan.2007.0116

Casati, R., & Varzi, A. (2020). Events. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Summer 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2020/entries/events.

Cohn, N., & Paczynski, M. (2019). The Neurophysiology of Event Processing in Language and Visual Events. In R. Truswell (Ed.), *The Oxford Handbook of Event Structure* (pp. 623–637). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199685318.013.26

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin and Review*, *24*(4), 1158–1170. https://doi.org/10.3758/S13423-016-1191-6

Craik, K. J. W. (1943). *The Nature of Explanation*. Cambridge University Press.

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychologica*, *125*(2), 129–143. https://doi.org/10.1016/j.actpsy.2006.07.004

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619. https://doi.org/10.1353/lan.1991.0021

Dowty, D. R. (1989). On the Semantic Content of the Notion of 'Thematic Role.' In G. Chierchia, B. H. Partee, & R. Turner (Eds.), *Properties, Types and Meaning* (Vol. 39, pp. 69–129). Springer Netherlands. https://doi.org/10.1007/978-94-009-2723-0_3

Ezzyat, Y., & Davachi, L. (2011). What Constitutes an Episode in Episodic Memory? *Psychological Science*, *22*(2), 243–252. https://doi.org/10.1177/0956797610393742

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating Verbs, Situation Schemas, and Thematic Role Concepts. *Journal of Memory and Language*, *44*(4), 516–547. https://doi.org/10.1006/jmla.2000.2728

Fillmore, C. J. (1968). The Case for Case. In E. Bach, & R. T. Harms (Eds.), *Universals in Linguistic Theory* (Vol. 2, pp. 1–90). Holt, Rinehart and Winston.

Fillmore, C. J., & Baker, C. (2009). A frames approach to semantic analysis. In B. Heine, & H. Narrog (Eds.), The Oxford handbook of linguistic analysis. Oxford University Press.

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. https://doi.org/10.1017/langcog.2014.20

Fuchs, T. (2017). *Ecology of the Brain: The phenomenology and biology of the embodied mind*. Oxford University Press. https://doi.org/10.1093/med/9780199646883.001.0001

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, *83*(1), 49–79. https://doi.org/10.1016/S0010-0277(01)00166-4

Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition*, *89*(1), B27–B41. https://doi.org/10.1016/S0010-0277(03)00069-6

Gerwien, J. (2011). A psycholinguistic approach to AT-structure analysis. In K. Spalek, & J. Domke (Eds.), *Sprachliche Variationen, Varietäten und Kontexte. Festschrift für Rainer Dietrich*. Stauffenburg.

Gerwien, J., & Flecken, M. (2016). *First things first? Top-down influences on event apprehension*. 2633–2638.

Gerwien, J., & von Stutterheim, C. (2022). 6. Describing motion events. In A. H. Jucker, & H. Hausendorf (Eds.), *Pragmatics of Space* (pp. 153–180). De Gruyter. https://doi.org/10.1515/9783110693713-006

Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. *Cognition*, *180*, 225–237. https://doi.org/10.1016/j.cognition.2018.07.008

Gerwien, J., & von Stutterheim, C. (2022). Describing motion events. In A. H. Jucker, & Heiko. Hausendorf (Eds.), *HoPs 14 Pragmatics of Space*. De Gruyter Mouton.

Gerwien, J., von Stutterheim, C., & Rummel, J. (2022). What is the interference in "verbal interference"? *Acta Psychologica*, *230*, 103774. https://doi.org/10.1016/j.actpsy.2022.103774

Glanemann, R., Zwitserlood, P., Bölte, J., & Dobel, C. (2016). Rapid apprehension of the coherence of action scenes. *Psychonomic Bulletin & Review*, *23*(5), 1566–1575. https://doi.org/10.3758/s13423-016-1004-y

Gold, D. A., Zacks, J. M., & Flores, S. (2017). Effects of cues to event segmentation on subsequent memory. *Cognitive Research: Principles and Implications*, *2*(1), 1. https://doi.org/10.1186/s41235-016-0043-2

Goldin-Meadow, S. (2003). *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language*. Psychology Press.

Goldman, A. I. (1970). *Theory of human action*. Princeton University Press.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. https://doi.org/10.1111/1467-9280.00255

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, *142*(3), 880–905. https://doi.org/10.1037/a0030045

Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thompson-Schill, S. L. (2012). The Effect of Object State-Changes on Event Processing: Do Objects Compete with Themselves? *Journal of Neuroscience*, *32*(17), 5795–5803. https://doi.org/10.1523/JNEUROSCI.6294-11.2012

Hindy, N. C., Solomon, S. H., Altmann, G. T. M., & Thompson-Schill, S. L. (2015). A Cortical Network for the Encoding of Object Change. *Cerebral Cortex*, *25*(4), 884–894. https://doi.org/10.1093/cercor/bht275

Horchak, O. V., & Garrido, M. V. (2021). Dropping bowling balls on tomatoes: Representations of object state-changes during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(5), 838–857. https://doi.org/10.1037/xlm0000980

Jeunehomme, O., & D'Argembeau, A. (2020). Event segmentation and the temporal compression of experience in episodic memory. *Psychological Research*, *84*(2), 481–490. https://doi.org/10.1007/s00426-018-1047-y

Johnson, C. R., Schwarzer-Petruck, M., Baker, C. F., Ellsworth, M., Ruppenhofer, J., & Fillmore, C. J. (2016). *FrameNet: Theory and practice*. International Computer Science Institute.

Johnson-Laird, J. P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., & Altmann, G. T. M. (2020). The influence of state change on object representations in language comprehension. *Memory & Cognition*, *48*(3), 390–399. https://doi.org/10.3758/s13421-019-00977-7

Kang, X., Joergensen, G. H., & Altmann, G. T. M. (2020). The activation of object-state representations during online language comprehension. *Acta Psychologica*, *210*, 103162. https://doi.org/10.1016/j.actpsy.2020.103162

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Klein, W. (1999). Wie sich das deutsche Perfekt zusammensetzt. *Zeitschrift für Literaturwissenschaft und Linguistik*, *29*(1), 52–85. https://doi.org/10.1007/BF03379170

Klein, W. (2010). On times and arguments. *Linguistics*, *48*(6). https://doi.org/10.1515/ling.2010.040

Kuperberg, G. R. (2021). Tea With Milk? A Hierarchical Generative Framework of Sequential Event Comprehension. *Topics in Cognitive Science*, *13*(1), 256–298. https://doi.org/10.1111/tops.12518

Lambert, M., von Stutterheim, C., Carroll, M., & Gerwien, J. (2022). Under the surface: A survey of principles of language use in advanced L2 speakers. *Language, Interaction and Acquisition*, *13*(1), 1–28. https://doi.org/10.1075/lia.21014.lam

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Löbner, S. (2021). Cascades. Goldman's Level-Generation, Multilevel Categorization of Action, and Multilevel Verb Semantics. In S. Löbner, T. Gamerschlag,

T. Kalenscher, M. Schrenk, & H. Zeevat (Eds.), *Concepts, Frames and Cascades in Semantics, Cognition and Ontology* (Vol. 7, pp. 263–307). Springer International Publishing. https://doi.org/10.1007/978-3-030-50200-3_13

Lupyan, G. (2012). Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00054

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, *33*(7), 1174–1184. https://doi.org/10.3758/BF03193221

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and Perception:* Harvard University Press. https://doi.org/10.4159/harvard.9780674421288

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28–38. https://doi.org/10.1037/h0035584

Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. Psychology Press.

Papafragou, A. (2015). The Representation of Events in Language and Cognition. In E. Margolis, & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 327–346). MIT Press.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*(1), 155–184. https://doi.org/10.1016/j.cognition.2008.02.007

Prystauka, Y. (2018). Comprehending Events on the Fly: Inhibition and Selection during Sentence Processing. *Master's Theses. 1206.*

Radvansky, G. A. (2012). Across the Event Horizon. *Current Directions in Psychological Science*, *21*(4), 269–272. https://doi.org/10.1177/0963721412451274

Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin & Review*, *26*(6), 1850–1869. https://doi.org/10.3758/s13423-019-01634-5

Sauppe, S., & Flecken, M. (2021). Speaking for seeing: Sentence structure guides visual event apprehension. *Cognition*, *206*, 104516. https://doi.org/10.1016/j.cognition.2020.104516

Solomon, S. H., Hindy, N. C., Altmann, G. T. M., & Thompson-Schill, S. L. (2015). Competition between Mutually Exclusive Object States in Event Comprehension. *Journal of Cognitive Neuroscience*, *27*(12), 2324–2338. https://doi.org/10.1162/jocn_a_00866

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human Brain Activity Time-Locked to Narrative Event Boundaries. *Psychological Science*, *18*(5), 449–455. https://doi.org/10.1111/j.1467-9280.2007.01920.x

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236–257. https://doi.org/10.1037/a0015631

Taylor, H., & Vickers, P. (2017). Conceptual fragmentation and the rise of eliminativism. *European Journal for Philosophy of Science*, *7*(1), 17–40. https://doi.org/10.1007/s13194-016-0136-2

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. https://doi.org/10.1016/j.jml.2010.02.006

Ünal, E., Ji, Y., & Papafragou, A. (2021). From Event Representation to Linguistic Meaning. *Topics in Cognitive Science*, *13*(1), 224–242. https://doi.org/10.1111/tops.12475

Van Dijk, T. A. (1995). On macrostructures, mental models, and other inventions: A brief personal history of the Kintsch-van Dijk theory. In *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 383–410).

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

von Stutterheim, C., Gerwien, J., Bouhaous, A., Carroll, M., & Lambert, M. (2020). What makes up a reportable event in a language? Motion events as an important test domain in linguistic typology. *Linguistics*, *58*(6), 1659–1700. https://doi.org/10.1515/ling-2020-0212

von Stutterheim, C., & Gerwien, J. (2023). Die Bedeutung sprachspezifischer Ereignisschemata für die Argumentstruktur. Ein Vergleich zwischen dem Ausdruck von Bewegungsereignissen im Deutschen und im Französischen. In J. Hartmann, & A. Wöllstein (Eds.), *Propositionale Argumente im Sprachvergleich | Propositional Arguments in Cross-Linguistic Research. Theorie und Empirie | Theoretical and Empirical Issues (Studien zur deutschen Sprache 84)*

Wilson, F., Papafragou, A., Bunger, A., & Trueswell, J. C. (2011). Rapid extraction of event participants in caused motion events. *Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 33, No. 33)*.

Wilson, V. A. D., Zuberbühler, K., & Bickel, B. (2022). The evolutionary origins of syntax: Event cognition in nonhuman primates. *Science Advances*, *8*(25), eabn8464. https://doi.org/10.1126/sciadv.abn8464

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253–265. https://doi.org/10.1002/WCS.104

Zacks, J. M. (2020). Event Perception and Memory. *Annual Review of Psychology*, *71*(1), 165–191. https://doi.org/10.1146/annurev-psych-010419-051101

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*(6), Article 6. https://doi.org/10.1038/88486

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

Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and Aging*, *21*(3), 466–482. https://doi.org/10.1037/0882-7974.21.3.466

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

Zuberbühler, K., & Bickel, B. (2022). Transition to language: From agent perception to event representation. *WIREs Cognitive Science*, *13*(6). https://doi.org/10.1002/wcs.1594