COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 43 (2019) 1–30

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12708

Keeping the Result in Sight and Mind: General Cognitive Principles and Language-Specific Influences in the Perception and Memory of Resultative Events

Maria Sakarias,[a] Monique Flecken[b]

[a]*Neurobiology of Language Department, Max Planck Institute for Psycholinguistics*
[b]*Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen*

Received 30 July 2017; received in revised form 3 August 2018; accepted 26 November 2018

Abstract

We study how people attend to and memorize endings of events that differ in the degree to which objects in them are affected by an action: *Resultative* events show objects that undergo a visually salient change in state during the course of the event (peeling a potato), and *non-resultative* events involve objects that undergo no, or only partial state change (stirring in a pan). We investigate general cognitive principles, and potential language-specific influences, in verbal and nonverbal event encoding and memory, across two experiments with Dutch and Estonian participants. Estonian marks a viewer's perspective on an event's result obligatorily via grammatical case on direct object nouns: Objects undergoing a partial/full change in state in an event are marked with partitive/accusative case, respectively. Therefore, we hypothesized increased saliency of object states and event results in Estonian speakers, as compared to speakers of Dutch. Findings show (a) a general cognitive principle of attending carefully to endings of resultative events, implying cognitive saliency of object states in event processing; (b) a language-specific boost on attention and memory of event results under verbal task demands in Estonian speakers. Results are discussed in relation to theories of event cognition, linguistic relativity, and thinking for speaking.

Keywords: Event cognition; Cross-linguistic analysis; Visual attention; Recognition memory; Causative events; Grammatical case; Thinking for speaking; Linguistic relativity

1. Introduction

Human beings segment the world around them into discrete units of dynamic action involving some sort of change over time; these units are what we call "events" (e.g.,

Correspondence should be sent to Monique Flecken, Neurobiology of Language Department, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH, Nijmegen, the Netherlands. E-mail: monique.flecken@mpi.nl

Newtson, 1973; Radvansky & Zacks, 2014; Shipley & Zacks, 2008; Zacks & Swallow, 2007). Event segmentation is a fundamental process in perception, operating on our predictions of what is going to happen next when tracing a sequence of actions. When we can no longer accurately predict what is coming, we update our event models and perceive an event boundary. Event boundaries are thus perceived during temporal windows in which a change along a specific dimension in an event accumulates, leading to maximal prediction error (e.g., Zacks et al., 2001). The changes accumulating in an event can be of different nature, for example, changes of spatial location of event participants as in a motion event (person crossing the street), or changes of state in event participants. The latter holds for causative events in which an inanimate patient-object is changed by an agent's actions, whereby the patient's qualitative properties are altered, for example, a person peeling a potato, in which, in the course of the event, the potato changes from unpeeled to peeled. This study explores how the ways in which such causative events can be conveyed linguistically impact event perception and memory.

When describing the potato-peeling event to another person in Dutch, for example, one is likely to use a simple past tense sentence such as *de man schilde een aardappel* ("the man peeled a potato"). Interestingly, using language the degree of focus on the event's boundary, in this case, the change in the potato from unpeeled to fully peeled, can be modulated. Through specific linguistic means, one can make explicit to what extent an event boundary was actually reached or not, during the specific time span for which one's assertion about the event holds. More interestingly even, the availability of means to make this type of information explicit varies cross-linguistically: While in some languages there are grammaticalized — and thus frequently used — means to specify the result of a causative action, in other languages this event dimension is often left unspecified. For instance, the Dutch example above does not make explicit whether the potato-peeling action has led to a full change in state of the potato involved, that is, a fully peeled potato. One can imagine a sentence continuation such as … *but whilst doing so, he received an emergency phone call and had to leave before finishing* which would cancel out the interpretation of the action as having led to a full result. A language like Estonian, on the other hand, provides grammatical means to mark whether an object in an event has changed substantially or not: Object nouns in transitive sentences are obligatorily marked for grammatical case, distinguishing between objects that have been affected only partially (partitive case, which would apply to a half-peeled potato, *naine kooris kartulit* "girl peeled potato-PART") or fully (accusative case, which would be used to describe a fully peeled potato after a peeling-action had been finished, *naine kooris kartuli* "girl peeled potato-ACC") (Ackerman & Moore, 2001; Kiparsky, 1998). Object case marking in Estonian thus provides information on the viewer's perspective on the result of an event: Has the result, that is, the object state change, been achieved partially or fully? Of course, in Dutch one *can* specify information on event results (using tense-aspect contrasts, for example, *was een aardappel aan het schillen* "was peeling" vs. *heeft een aardappel geschild* "has peeled a potato"), but, in contrast to a language like Estonian where object case marking is obligatory in each and every sentence, this event perspective is often simply underspecified in an event description.

The intriguing questions we pose here are how are (prospective) results of causative events visually processed and memorized in languages that mark this dimension overtly on a habitual basis, compared to languages which encode it optionally and less frequently? To what extent does explicit verbalization of the events influence the cross-linguistic perception and memory of their results? One hypothesis is that this dimension may be universally salient, given the importance attributed to the detection of event boundaries in action perception (see, e.g., Radvansky & Zacks, 2011). In addition, and of particular relevance to the semantic domain of caused action and transitivity, it has been proposed that events may be represented in terms of intersecting representations of the objects in them. Object representations, which include information on their state as either unchanged or changed by the action in an event, carry spatial-temporal properties on which entire event representations are built (see Altmann & Mirkovic, 2009; Altmann, 2017; Hindy, Altmann, Kalenik, & Thompson-Schill, 2012; Solomon, Hindy, Altmann, & Thompson-Schill, 2015). This means that people will automatically encode and store information on the actions inflicted upon objects and the objects' specific (end) states in order to form their representation of an event. However, taking into account differences in the extent to which object states are marked linguistically, it may well be that our perception and memory of event results are susceptible to linguistic modulation and cross-linguistic variation. The alternative hypothesis underlying the present investigation is, therefore, that habitual practice in the linguistic encoding of object affectedness and their states as representing either a partial or full event result may lead to stronger cognitive biases toward this event dimension in general: The degree of automaticity and habituation associated with grammaticalized distinctions in particular may lead to linguistic relativity effects (cf. Gumperz & Levinson, 1996; Lucy, 1992), which, in this case, would be reflected in a task-independent prominence of event results in visual attention and memory of Estonian speakers. As a second alternative, grammar could influence attention processes and memory, but mainly in contexts when those grammatical structures have to be retrieved and accessed online, during language planning and formulation when describing events (during the process of "thinking for speaking," Slobin, 1996, 2003), but not when perceiving events generally. The present cross-linguistic comparison will thus shed important light on general cognitive tendencies, as well as language-specific biases, concerning the saliency of event boundaries, focusing on object state change in particular.

We use an experimental design that taps into both online event encoding (visual attention allocation) and event memory after encoding under varied task demands (see e.g., Bunger, Skordos, Trueswell, & Papafragou, 2016; Flecken, von Stutterheim, & Carroll, 2014; Papafragou, Hulbert, & Trueswell, 2008; Soroli & Hickmann, 2010). We record eye movements in Dutch and Estonian participants, while they are visually inspecting live-recorded dynamic event stimuli and analyze their attention to the action and the patient-object affected by it. After encoding, we administered a surprise recognition memory task, to investigate whether, and how detailed, information on an event's result was committed to one's memory of an event cross-linguistically. By including two experimental task conditions for visual scene encoding, one requiring overt verbalization (event description), the other involving a non-verbal distracter task, we can compare the extent to which grammar influences event processing task-specifically (i.e., while preparing to speak about an event) or -independently, for the hitherto largely unexplored semantic domain of object state change and event results.

1.1. Event cognition cross-linguistically

Prior cross-linguistic analyses of event processing mainly focused on the semantic domain of motion events. The bulk of work studied cross-linguistic variation in lexical-semantics (but see Pavlenko & Volynsky, 2015): It has been investigated whether speakers of verb- and satellite-framed languages differ in how they (learn to) perceive, describe, categorize, and memorize motion (e.g., Bohnemeyer et al., 2007; Carroll, Weimar, Flecken, Lambert, & von Stutterheim, 2012; Filipović, 2011; Gennari, Sloman, Malt, & Fitch, 2002; Han & Cadierno, 2010; Kersten et al., 2010; Ji, Hendriks, & Hickmann, 2011; Papafragou, Massey, & Gleitman, 2002; Papafragou et al., 2008). Typically, the motion verbs in a satellite-framed language encode information on the manner of motion (e.g., *to tiptoe, to crawl*), whereas most motion verbs in a verb-framed language contain path-information (on the trajectory or goal of motion, for example, *to enter, to approach*; Talmy, 1985, 2000). These typological differences give rise to differences in the saliency that viewers attribute to manner and path-elements of a motion event. For example, speakers of satellite-framed languages typically mention the manner of motion, and they also allocate more visual attention to manner when watching and describing event videos, compared to speakers of verb-framed languages (e.g., Papafragou et al., 2008; Soroli & Hickmann, 2010). There is, however, not much evidence that such effects go beyond "thinking for speaking," that is, beyond the cognitive processes we engage in when preparing for verbalization (e.g., Bunger et al., 2016; Finkbeiner, Nicol, Greth, & Nakamura, 2002; Gennari et al., 2002; Montero-Melis et al., 2017; Papafragou et al., 2002, 2008; Trueswell & Papafragou, 2010). Montero-Melis et al. (2017) plausibly suggest that this may be caused by the large variability in motion event description, both within speakers of the same typological cluster, as well as across clusters. To top it all, the typological difference mainly concerns the saliency of manner, which is considered "peripheral" information in a motion event, whereas the path represents the universal core of a motion event (Talmy, 2000). In sum, differences in motion event framing typology cannot reliably be linked to cognitive differences beyond speaking contexts (see, e.g., Bunger et al., 2016; Papafragou et al., 2008).

Another line of research has studied linguistic variation in the realm of grammar, that is, grammatical aspect systems, and its influence on motion cognition. When viewing motion events in which entities were on their way toward a goal without actually reaching it (e.g., a woman walking toward a bus stop within a reasonable distance), speakers of languages with progressive aspect (e.g., English) focused mainly on the ongoing action, rather than the goal (e.g., Athanasopoulos & Bylund, 2013; Flecken et al., 2014; Flecken, Athanasopoulos, Kuipers, & Thierry, 2015; von Stutterheim & Nüse, 2003). This pattern differed from the one found in speakers of languages without grammatical aspect (e.g., Swedish, German), who, in turn, showed a bias toward attending to and mentioning the goal. The evidence for cross-linguistic differences in this domain also spans attention and categorization patterns during non-speaking tasks, suggesting a task-independent influence of grammar on motion cognition (e.g., Athanasapoulos et al., 2015; Flecken, Athanasopoulos, et al., 2015).

The domain of causative events has been studied in this field as well, though to a lesser extent (e.g., Ji, Hendriks, & Hickmann, 2011; Wolff & Ventura, 2009). Wolff and Ventura (2009), for example, looked at English and Korean. In Korean, inanimate entities cannot be mapped onto the role of causer (agent) in an event; a sentence such as "the rock broke the windshield" would be unacceptable. In an event segmentation task, speakers of Korean were less likely than speakers of English to consider such instances an event in its own right; they preferred to view it as part of a causal chain initiated by a human agent (e.g., a boy threw a rock, which broke the windshield). Another study focused on intentional versus accidental causative events and studied the extent to which variation in the encoding of agenthood influences event memory (Fausey & Boroditsky, 2011). Spanish, in contrast to English, allows *pro*-drop (omitting the subject of a sentence, referring to the agent in an event, for example). In a description task and subsequent memory experiment, Spanish speakers mentioned and remembered agents of events less frequently compared to speakers of English.

Overall, the experimental evidence concerning the broader issue of general cognitive and language-specific biases in event cognition is mixed: Researchers have compared different languages, studying different event types with different experimental paradigms, leading to a heterogeneous picture. Most of the work has targeted the domain of motion events; the work on another core event type, that is, causative events (involving an agent performing action on an object), mainly centers on agentivity, leaving the dimension of object states and event results largely unexplored.

2. The present study

Two groups of native speakers of Estonian or Dutch viewed short (3-second) videoclips of causative events, while their eye movements were being recorded. We analyze their attention to the action and the object in the event during the final phases of the event's unfolding (video endings). After the encoding phase, participants performed a surprise recognition memory task that tested their memory of video endings, that is, whether the action they had seen in the video earlier had been finished or not. Participants were randomly assigned to two encoding conditions. In one condition, participants described each video in one sentence after it had stopped playing (verbal encoding condition). In the other condition, they inspected the videos silently while performing a distracter task that instructed them to pay attention to a continuous background sound and remember the content of those videos in which an additional sound cue (a beep) was played (non-verbal encoding condition, following Flecken et al., 2014). In both conditions, two additional neuropsychological experiments (Digit Span and Corsi-Blocks tapping tasks) assessed visual-spatial and verbal working memory capacity in order to control for general differences in memory across populations. The manipulation of task demands, plus the cross-linguistic contrast between Dutch and Estonian, are used as windows onto general cognitive biases in, and potential linguistic modulation of, the perception and memory of event results.

Estonian transitive event descriptions differ from Dutch ones in that information on the result of an event, in particular, the degree to which an object is changed in state by an action, is expressed grammatically in the opposition between accusative and partitive case-marking of nouns referring to direct objects (e.g., Kiparsky, 1998; Lees, 2004; Tamm, 2004, 2007). Accusative case is used to mark direct object nouns that refer to quantitatively bound objects, affected in their totality by actions that have been completed. Partitive case refers to partially affected objects in events. Object case marking in Estonian thus conveys information on whether a speaker viewed the event as having concluded with a partial (object changed in state partially) or a full (object changed in state fully) result. Importantly, taking a viewpoint on an event's result is obligatory in every event description that contains reference to a direct object. Dutch, the other language under investigation, does not mark grammatical case, nor does it provide other obligatory means for expressing a speaker's viewpoint on changes in state of objects and event results.[1] Thus, by investigating verbal and non-verbal event encoding in Estonian and Dutch participants, we explore, first, how visual processing for speaking generally differs from non-verbal scene perception, that is, do verbalization requirements generally enhance attention to and memory of how events end? This would be demonstrated by a main effect of encoding condition on visual attention and event memory, and no interaction with language background.

Second, we ask whether there is evidence for a "thinking for speaking" effect: Estonian speakers, given grammatical marking of direct objects as partially or fully changed in state, are likely to show a language-specific boost on perception and memory of event results in the verbal condition; this would be demonstrated by a language by encoding condition interaction. If, however, we find a main effect of language on our dependent variables, that is, a similar language-specific boost in the non-verbal condition, this will be evidence for a linguistic relativity effect. It would show that language-specific grammatical requirements can lead to general, task-independent cognitive biases (see e.g., Boroditsky, Fuhrman, & McCormick, 2011; Casasanto, 2008; Gumperz & Levinson, 1996; Imai, Schalk, Saalbach, & Okada, 2014; Lucy, 1992; Wolff & Holmes, 2011).

Importantly, the factor "event type" was also included in our design: *Resultative events* involved objects that underwent a visually salient change in state within the duration of the videos (e.g., *cut a circle, peel a potato*). By manipulating whether at video offset the action was depicted as ceased or still ongoing, we expected to trigger case-marking alternation in Estonian descriptions: Resultative event videos ending after 3-s with a still ongoing action should be predominantly marked with partitive case (e.g., "kartuli-t", potato.PART), as the event concluded with a partial result. Ceased resultative events should elicit accusative case (e.g., "kartuli," potato.ACC, referring to a fully peeled potato). *Non-resultative events* included actions in which objects did not visibly undergo substantial change throughout the time span of the videos, for example, *stir a bowl of* *soup, measure a box* (inherently no object state change), and for example, *grate cheese, staple a pile of papers* (object is affected, but, during the time window depicted in the videos, the change was partial, for example, grate some cheese from a large chunk of parmesan). Here, partitive case is the most likely choice for both ongoing and ceased action videos (e.g., "raamatu-t", book.PART). This manipulation allows an in-depth investigation of "thinking for speaking" processes. In particular, does a potential verbal boost on attention and memory in Estonian apply to *both* event types, or only to resultative events? First, a "thinking for speaking" effect could relate to the processing of resultative events only: Estonian speakers would encode video endings more attentively than Dutch speakers, because this information is critical for online linguistic retrieval of the form of the direct object noun. This is only the case for resultative events, given that, whether the action is ceased or ongoing has consequences for the visual state of the objects in them; it means either fully or partially affected objects, which drives the choice between partitive and accusative case in Estonian. The ceased/ongoing manipulation does not affect the objects' visual states in the same way for non-resultative events. Estonian speakers should thus inspect video endings of resultative events carefully to get information on the event's result and to formulate the right noun-ending. This scenario would surface as a three-way interaction of condition, language, and event type in our analyses of attention and memory. Interestingly, one might also argue that the use of the *accusative* case specifically, marking a fully changed object, drives a potential bias toward event results in Estonian participants. Only in the case of full object state change, an event boundary is *reached* which might induce extra attention. To explore if a "thinking for speaking" effect is linked to the specific object-case marker selected, we run additional analyses with "case" as a binomial predictor of fixation and memory patterns.

Alternatively, a "thinking for speaking" effect could apply to *both* event types: Estonian speakers are required by their grammar to mark a viewpoint on an event's result, regardless of whether *in the moment* they are faced with the selection of partitive or accusative case, and regardless of *which specific form* is selected. This is not the case for Dutch participants. The additional processing demands associated with providing an explicit viewpoint on the result in an event could plausibly drive an attentional and memory boost in Estonian participants regardless of event type.

Lastly, with the manipulation of event type, we can find out whether resultative events, compared to non-resultative events, trigger more attention and are represented more robustly in memory, given the variability in terms of object states depicted at the end of the event, independent of language-specific demands. This would surface as a main effect of event type on our dependent variables.

In sum, this study contributes to our understanding of cross-linguistic influences on event perception, description, and memory in various ways: First, the specific cross-linguistic contrast allows uncovering potential processing differences induced by grammatical categories, different from most previous work. Importantly, the linguistic variability applies to the encoding of an event dimension which is attributed universal importance in event cognition, namely event boundaries (the way in which they end) in the form of objects that underwent more or less change by the action in an event (cf. Altmann, 2017;

Kemmerer, 2012; Radvansky & Zacks, 2014). Second, by including a manipulation of event type controlling for case alternation in Estonian, we can provide a detailed picture of thinking for speaking processes. Third, this manipulation will advance our understanding of the saliency of event results and object state change generally. Fourth, the combination of online and offline measures (visual attention and memory) while varying encoding condition (verbal and non-verbal) across tasks will give us a more complete picture of event processing in two languages.

3. Methods

3.1. Participants

In total, 58 native speakers of Estonian were tested at the University of Tartu, Estonia, 33 of which were tested in the verbal encoding condition, and 25 of them in the non-verbal encoding condition.[2] Ten participants in the verbal encoding condition had to be excluded due to technical error or failure to fulfil task requirements.[3] The final sample consisted of 23 participants (mean age 22.57, *SD* = 3.47, *n* = 17 female). In the non-verbal encoding condition, two participants were excluded[4] (final sample of 23 participants with a mean age of 24.39 years [*SD* = 3.96], *n* = 16 female). The Dutch group included 50 native speakers of Dutch, recruited from the participant pool of the Max Planck Institute for Psycholinguistics, 26 of which were tested in the verbal encoding condition, 24 in the non-verbal encoding condition. In the verbal task, data from two participants were excluded due to technical error,[5] leaving a final group of 24 participants with a mean age of 22.46 (*SD* = 3.32, *n* = 17 female). In the non-verbal condition, two participants were excluded[6] (mean age of the final sample of 24 participants was 21.68 years [*SD* = 2.40], *n* = 17 female). All participants were right-handed, had normal or corrected-to-normal vision, and had no neurological or psychological disorders. All participants gave written consent to take part in the experiment and received payment for participation.

3.2. Materials[7]

Video-clips were recorded at the Max Planck Institute for Psycholinguistics for this study. They showed four different actors (three female, one male) performing every-day actions on various objects. The actions were selected and performed ensuring maximal spatial separation of the two main elements of the event (agent and action/object, allowing the definition of two Areas of Interest for later eye tracking analyses, see Fig. 1 and a full list of items in Appendix A). All videos were filmed against a white background with no distracting items. Videos were cut to last 3,000 ms and they showed *Resultative events* (*N* = 18), *Non-resultative events* (*N* = 18) and fillers (*N* = 18). Resultative events included action on a single, specific object, leading to a visually salient change in state during the course of the event. The visual states of the objects in them made it plausible that a full change in state, and thus a full event result, would be achieved by the end of the videos (e.g., folding an airplane from a sheet of paper, cutting a circle out of paper, peeling a banana). Non-resultative events included action on a single object as well, but the videos did not show a visually salient change in state of the object during the time window cut, and so the events were not likely to produce a full result in the given time span[8] (e.g., rub a knife with a cloth, polish glasses). Fillers included one-participant events (*n* = 5; e.g., person yawning), two-animate participant events (*n* = 6; e.g., man giving a book to a woman), and one-animate/one-inanimate participant events (*n* = 8) with no spatial separation between agent and action/object (e.g., person talking on the phone).

We manipulated the phase of the event depicted at the end of each video: Half of the stimuli ended after 3 s showing the action as still ongoing; in the other half, after 3 s the action was ceased by the agent. Each video was cut twice, one version showing ongoing action, the other showing ceased action. For resultative events, ongoing versions implied that the actions had only produced a partial result at the end of the video, whereas their ceased counterparts showed the achievement of a full result (actor cuts circle completely, puts down scissors and withdraws hands from the object). In non-resultative events, the actions depicted led to none or only a partial result; the objects involved in the actions did not undergo a visually salient change in state in either version (ongoing version: stirring still ongoing by video offset; ceased version: agent stops stirring, puts down the spoon next to the plate, and withdraws hands from the object). The videos revealed whether or not an action was going to be ceased around 300 ms before video offset. At this point in time, in ceased versions, culmination of the event would become clearly visible. This variation was included to trigger case alternation in Estonian verbal descriptions, allowing us to address the questions outlined in relation to the factor “event type” in section 2 above. In addition, it should generally increase the attention paid to video endings in both groups. As we had no a priori hypotheses concerning attention and memory to ceased versus ongoing versions of the events, this distinction is not included in any of the analyses (in all analyses ceased and ongoing trials are collapsed).

Four stimulus lists of 54 video clips were constructed, such that two lists included the ceased version of an event, and two included its ongoing version. The number of ceased

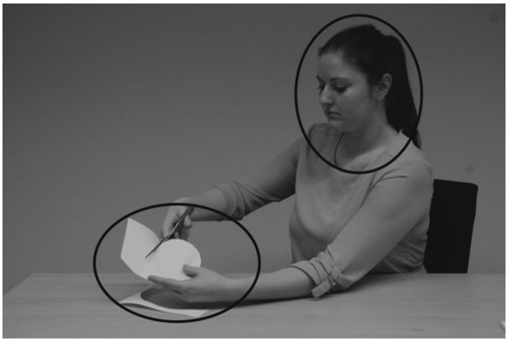


Fig. 1. Example of a stimulus (screenshot of the item *cut a circle*). Areas of Interest (not visible to participants) are marked by ellipses (Agent: area of head and shoulders of the actor; Action: area of action and the object).

and ongoing items was the same in each list and each participant saw each event once. Within each list, we pseudorandomized the position of the actor in the video (left or right) and the agent performing the action, such that all four actors appeared an equal number of times.

Materials for the event recognition task included screenshots of one of the last frames of the video clips, showing the action clearly as still ongoing or as ceased.

3.3. Procedure

Participants signed written consent and filled out a language background questionnaire. They were seated approximately 60 cm from the remote SMI RED250-m eye-tracker (SensoMotoric Instruments). The display resolution of the laptop was 1,920 × 1,080, and the eye-tracker recorded both eyes at 250 Hz. Participants performed four tasks in the following order: (a) Video encoding task; (b) Corsi-Blocks tapping task; (c) Event recognition task; (d) Digit span task. All experiments were programmed in Presentation (NeuroBehavioral Systems) and run on the SMI eye tracker. Eye movements were only recorded during the Video encoding tasks. An experimental session lasted approximately 35 min.

3.3.1. Verbal encoding task

Participants were instructed (in their native language) to watch each video until the end (which was signaled by a beep) and to describe the event depicted in one sentence, answering the question “What happened in the video clip?” (“Wat gebeurde er in de video?”, “Mis videos juhtus?”). The spacebar was used to proceed to the next video. Responses were recorded by an external microphone. Each video disappeared from the screen after playing, meaning that participants were looking at an empty screen while speaking. Participants first carried out two practise trials during which they could ask clarification questions. Participants did not receive feedback regarding sentence constructions used. Then, a semi-automatic 5-point calibration (controlled by SMI Eye View), which was repeated after every 12 trials, initiated the experiment. The task lasted approximately 10 min.

3.3.2. Non-verbal encoding task

Participants were engaged in a non-verbal sound-cue detection task (adapted from Flecken et al., 2014). They were asked to watch the videos in silence while listening to the sound of ocean waves being played continuously in the background. Randomly during some of the video clips, an additional loud beep would be played. Participants’ task was to remember during which of the videos this beep was played. Videos were presented in blocks, playing automatically one after another with 3,500 ms in between (2,000 ms white screen; 1,500 ms fixation cross). After a block of six video clips, participants saw a screenshot of one of them and had to respond to the question “Did you hear a beep during this video?”, which was written above the screenshot in the participant’s native language. After pressing a button corresponding to a “yes” or “no” response, the experiment continued with a new block of six video clips (a re-calibration was conducted every two blocks). Beeps were played during the midpoint of each video. Crucially, they only occurred during filler video clips to avoid drawing extra attention to and familiarization with the critical items. The number of beeps within a block alternated randomly between 1 and 3; in total, 10 response trials were recorded.

The task was designed to keep participants engaged and to focus their attention on the video clips without biasing them toward any of the event elements. Previous non-verbal paradigms involved instructing participants to inspect scenes carefully for an upcoming memory task which must have biased their attention to all aspects of the scenes, including details which would typically not be focused on (e.g., Papafragou et al., 2008). Moreover, under overt memorization instructions the use of inner speech and verbalization strategies cannot be excluded.

3.3.3. Corsi blocks tapping task

A mouse-based version of the Corsi Blocks tapping task was administered (Cognitive Experiments III v3, www.neurobs.com), measuring visual-spatial working memory (Kessels, Van Zandvoort, Postma, Kappelle, & De Haan, 2000). Participants were shown nine blue rectangles on a gray background, which turned red one by one. The task was to memorize the order in which the rectangles changed color, and to recreate the pattern after each trial by clicking on the relevant rectangles with the computer mouse. The test started with a trial length of three rectangles and increased in a 1:2 staircase method (following Woods et al., 2011), where a single correct response increased the length of the subsequent list by one rectangle and two incorrect responses reduced the length by one rectangle. The task ended after the participant had completed 14 trials. The duration of the task was kept constant by the experimenter who stopped the task after 3.5–4.0 min.

3.3.4. Recognition memory task

We used a two-alternative forced-choice task to test participants' memory of the endings of the previously encoded videos. Participants saw two screenshots side by side: One of them showed the actual ending of the video (e.g., ceased version: woman with an open can in front of her) while the other screenshot showed the other version (ongoing version, woman engaged in opening of the can, see Fig. 2). Participants were asked to press a button left or right (Q or P) on the keyboard, indicating which screenshot showed the final frame(s) of the video they had seen before, as fast as possible. The actor's identity and the position of the object on the screen was the same as during the encoding phase; the order of presentation of the items was also identical to the encoding task. In the recognition memory task, again four pseudo-randomized lists were used to vary the order of conditions/items.

3.3.5. Digit Span task

Finally, participants completed the forward Digit Span task (Cognitive Experiments III v3, www.neurobs.com), tapping into verbal working memory capacity. In each trial, a series of digits was presented one by one in the center of a white screen. Participants had to memorize the digits and type them in the correct sequence with the keyboard. List length of the trials started from 3 and increased in a 1:2 staircase method until 14 trials had been presented.

4. Data pre-processing, coding, and analyses

4.1. Control data: Corsi-blocks, digit span, non-verbal sound-cue recognition task

The method of obtaining a participant score for the Corsi Blocks and Digit Span control tasks was adopted from Woods et al. (2011), using a mean span (MS) metric as the most reliable and precise measure for quantifying the results of neuropsychological tasks. The MS baseline was set at 2.5 (0.5 digits less than the initial list length), and the score was calculated by adding the baseline to the rate of accuracy at each list length (see Woods et al., 2011).

Accuracy scores for sound-cue recognition in the non-verbal task were analyzed as well.

4.2. Event description data

Audio files were transcribed and coded for object case-marking in Estonian. Dutch transcriptions were inspected for use of linguistic means to mark the specific degree of object state change in the event, that is, systematic variation in tense-aspect (use of present perfect to mark event completion and use of progressive aspect to mark ongoingness, which has consequences for one's conception of the degree of object affectedness).[9] In addition, data were coded for use of linguistic means that marked a specific degree of state change in an object (e.g., cut paper in half, break chocolate into pieces). Coding was carried out by two coders independently. Discrepancies between the coders existed only on a very low number of trials (about 5% of all trials), and they were resolved after discussion.

4.3. Eye movement data

Two identically sized and spatially distant elliptical areas of interest (AoI) were defined for each stimulus after all data had been collected: One AoI included the head

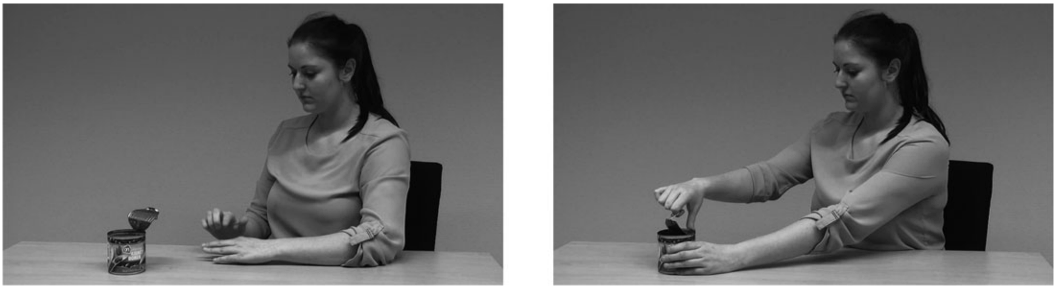


Fig. 2. Screenshot of a trial in the recognition memory task.

and shoulders of the actor (Agent), and the other AoI included the region of the action and the object (Action), encompassing the agent's hands as well as the object and the instrument fully (see Fig. 1).[10] The size of the two AoIs was kept constant across all stimuli. Fixations in these two AoIs were computed for the entire time that the videos were playing with SMI BeGaze software (SensoMotoric Instruments). During recording, Presentation software sent timestamps to the eye-tracker, marking stimulus onset and stimulus offset for each trial.

For plotting (see Appendix B), fixation reports were preprocessed in R (version 3.2.3), using a script which detected for each participant and each trial whether a fixation fell into a particular AoI in successive 50 ms bins (5,000 ms in total). Fixations were aggregated across participants for each AoI and time bin; data are presented as the proportion of fixations in a particular AoI during a given time bin. Plots also include fixations outside of both areas of interest.

The analyses focused only on looks in the Action AoI, given that our hypothesis was based on differential degrees of attention to an event's result, and the critical information is contained in this specific spatial region of the stimuli. In line with our hypothesis, we only focused on a subpart of the overall time course, namely the final phases of each event's unfolding (video endings), during which it became clear whether the action would finish, with potential consequences for the state of the object, or not. We computed the total duration of all fixations for each participant and each trial in a time window spanning 600 ms in total[11] around video offset. We then computed the log-transformed odds ratio of looking time in the Action area of interest for each item and participant in this window. A mixed effect linear regression model was used to predict the probability of fixations in the Action AoI on the basis of the predictors Condition (Non-verbal/Verbal), Language (Dutch/Estonian), and Event type (Non-resultative/Resultative). Predictors were effect coded (Condition "Nonverbal," Language "Dutch," and Event type "Non-resultative" were coded -1; the other levels were coded as 1). The model included random intercepts for participants and items, as well as by participant random slopes for the within-subject factor Event type, and by item random slopes for Condition.

In addition, we conducted an exploratory analysis on the Estonian verbal data to assess the extent to which the use of a specific case marker (accusative vs. partitive) influenced gaze behavior. We analyzed the same dependent variable as above, under the dependency of the binomial predictor Case (1 = accusative case; 0 = partitive case or other). The random effects structure of the model included random intercepts for participants and items, and a by-participant random slope for Case.

4.4. Recognition memory data

The analyses focused on accuracy of recognizing the result of the event: Participants had to select the picture showing the correct video ending and corresponding object state. We plotted the proportion of accurate responses for all items for each Condition, Language, and Event type. Binomial response data were analyzed with mixed effect logistic regression models, including the predictors Condition (Verbal/Non-verbal), Language (Dutch/Estonian), and Event type (Resultative/Non-resultative) (all factors were effect coded). Again, the model included random intercepts for participants and items, as well as a by-participant random slope for Event type, and a by-item random slope for Condition.

In addition, we conducted an exploratory analysis on data from the Estonian verbal encoding condition, asking whether the use of a specific case marker (accusative vs. partitive) could predict one's memory of an event's result. The random effects structure included random intercepts for participants and items, and a by-participant random slope for Case.

5. Results

5.1. Control data: Corsi Blocks-tapping task

Mean scores for the Corsi Blocks task are shown in Fig. 3.[12] A two-way ANOVA of Condition by Language showed no main effects of Condition ($F(1,87) = 0.003$, $p = .956$ ns), Language ($F(1,87) = .519$, $p = .473$ ns), and a non-significant trend for an interaction between the two factors ($F(1,87) = 3.293$, $p = .073$ ns). Performance on this task suggests that spatial working memory capacity did not differ in the four samples tested.

5.2. Control data: Digit Span task

Mean scores for the Digit Span task are presented in Fig. 4 below. A two-way ANOVA of Condition by Language showed no main effects of Condition ($F(1,88) = 0.594$,

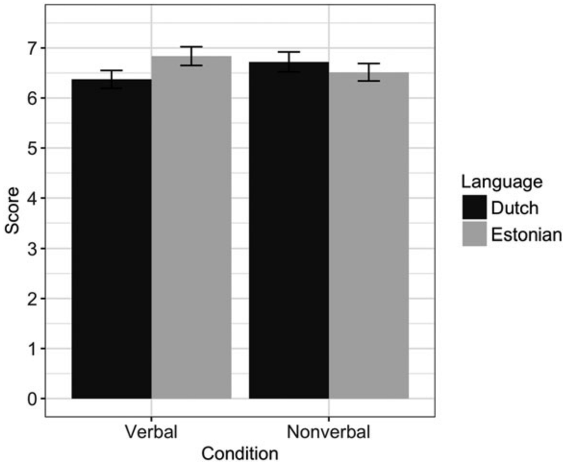


Fig. 3. Scores on the Corsi Blocks tapping task (error bars indicate ±*SE*).

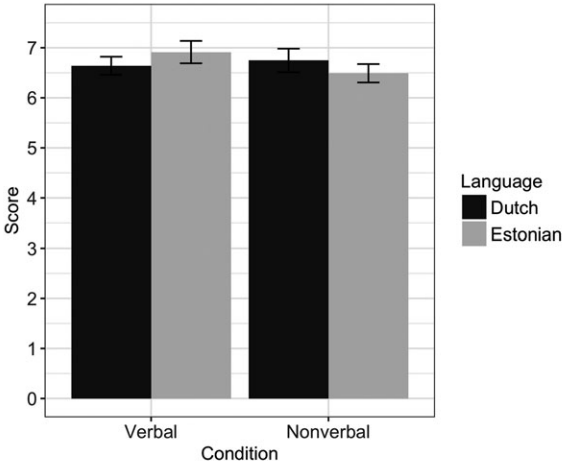


Fig. 4. Scores on the Digit Span task (error bars indicate ±*SE*).

p = .443 ns), Language (F(1,88) = 0.004, p = .948 ns), and no significant interaction (F(1,88) = 1.645, p = .203 ns). Verbal working memory capacity as measured by this task did not differ between groups.

5.3. Non-verbal encoding task: Sound-cue recognition accuracy

All participants demonstrated highly accurate performance on the sound-cue detection task: Mean scores for correctly identifying the scenes during which the sound cue had occurred was 91.3% in Estonian (*SD* = 28.24) and 93.3% in Dutch (*SD* = 25). A mixed effect logistic regression model on the binomial-dependent variable accuracy showed no significant effect of the factor Language (β = −0.145, *SE* = 0.181, p = .423 ns). This suggests that both groups understood the distracter task and paid attention to it during non-verbal scene encoding.

5.4. Verbal encoding task: Event descriptions

Table 1 below gives absolute and relative frequencies of partitive and accusative object case-marking in the Estonian descriptions.

A mixed effect logistic regression model on the binomial dependent variable “case” (1 = accusative, 0 = partitive), testing effects of the effect coded factors “Event type” and “End state” (ceased/ongoing), showed main effects of both factors (Intercept: β = −2.027, *SE* = 0.266, z = −7.609; Event type: β = 1.213, *SE* = 0.206, z = 5.887, p < .001; End state: β = 1.072, *SE* = 0.151, z = 7.123, p < .001), and an interaction (β = 0.349, *SE* = 0.147, z = 2.369, p < .05). The highest proportion of accusative case was found in descriptions of ceased resultative events.

Table 1
Case-marking in Estonian event descriptions (PART = partitive case, ACC = accusative case)

	Ongoing	Ceased	% of Total
Resultative events (N = 414)			
PART	176	79	61.6%
ACC	30	124	37.2%
n/a[13]	1	4	1.2%
Non-resultative events (N = 414)			
PART	188	165	85.3%
ACC	6	20	6.3%
n/a	14	21	8.4%

Participants were generally sensitive to our stimulus manipulations. However, descriptions of resultative events showed case alternation (object nouns were marked with accusative case in 37.2% of the cases, and the vast majority of these were elicited with *ceased* action videos; partitive case was used in 61.6% of resultative event descriptions), the predominant grammatical case in descriptions of non-resultative events was a partitive case. Although the accusative was thus most often used in events concluding with a full result (ceased resultative events), not *all* of these trials contained this case marker, showing that our manipulation did not fully constrain participants' conceptualization of the events and their subsequent linguistic choices. Note that, even when a resultative event concludes with a ceased action, for most of the time depicted the event in the video is in fact ongoing (and thus does not show a full result). When a speaker decides to focus on the time span prior to the event's final conclusion in his or her description, use of partitive case is grammatical. This is a possibility given that there were no instructions that forced participants to await the final frames of the videos before initiating speech preparation processes (note, though, that they were instructed to withhold speaking until a beep had sounded at video offset).

In the Dutch data, tense/aspect contrasts were not systematically used to mark the distinction between resultative, non-resultative and/or ceased, ongoing events: Descriptions were in present tense exclusively, with the exception of two trials in which past tense was used. Use of progressive aspect was negligible (only in 2.8% of the trials) and there was no use of the present perfect. In addition, we coded descriptions in both languages for linguistic ways of marking the degree of state change of an object, for example, resultative particles or prepositional phrases (PP) (e.g., cut an apple in half, break chocolate into pieces, pour a glass full). The question was whether Dutch speakers used these means more than Estonian participants, which would imply frequent marking of a viewpoint on an event's result in this group as well. Resultative particles/PPs were used in 9.15% of the Dutch data (15.78% of resultative and 2.55% of non-resultative event descriptions contained such forms), and in 12.35% of the Estonian data (23.30% of resultative and 1.45% of non-resultative descriptions contained these forms). A mixed effect logistic regression model, including the (effect-coded) factors "Language" and "Event type," and their interaction showed a main effect of Event type only (Intercept: $\beta = -10.778$, $SE = 1.754$, $z = -6.146$; Language: $\beta = -0.155$, $SE = 0.473$, $z = -0.327$

ns; Event type: $\beta = 6.8589$, $SE = 1.3307$, $z = 5.15$, $p < .001$; Language*Event type: $\beta = -0.3202$, $SE = 0.469$, $z = -0.682$ ns): In both groups, information on object state change was marked mainly in resultative event descriptions.

5.5. Eye movement data

For the interested reader, Appendix B contains line plots showing the proportion of fixations in the two Areas of Interest over time, for each Language, by Condition and Event type.

Fig. 5 below shows the proportion of looking time in the Action AoI during the predefined analysis window (final event phase) for each Condition separately.

A mixed effect linear regression model on logit transformed looking time in the Action AoI showed significant main effects of Condition and Event type, an interaction between Condition and Language, and a non-significant trend for a Language by Event type interaction (see Table 2).

The main effect of Condition shows that, overall, the Action AoI was fixated more in the Verbal than the Non-verbal encoding condition. There was also an overall higher likelihood of fixating the Action AoI in Resultative compared to Non-resultative events (main effect of Event type). The Condition by Language interaction is driven by a larger (significant) language effect in the Verbal than the Non-verbal condition (the analysis in fact only shows a trend for a language effect in the Non-verbal condition): In the Verbal encoding condition, Estonian participants have a higher likelihood of fixating the Action compared to Dutch participants. Their Action fixation probability drops in the Non-verbal encoding condition, in which the language patterns nearly reverse. Thus, Estonian

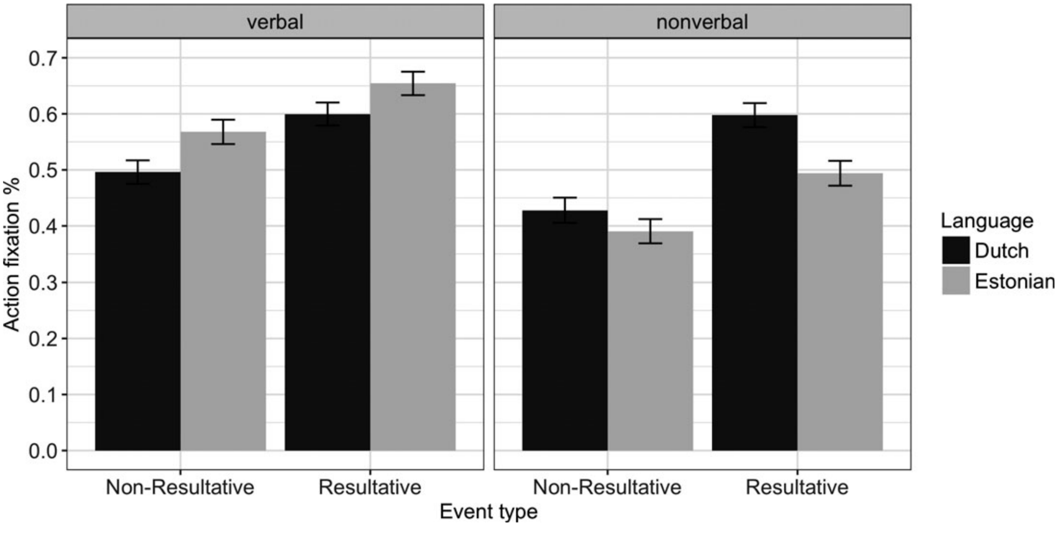


Fig. 5. Proportion of looking time in the Action AoI during final event phases, plotted for each Condition separately, by Language and Event type (error bars show ±*SE*).

Table 2
Output of the model on action fixation probability (lmer (logodds action fixation ~ condition * language * event type – language + (1 + event type|participant) + (1 + condition|item))

Fixed Effect	*β* Estimate	*SE*	*t*-Value
Intercept	0.317	0.252	1.261
Condition	0.649	0.210	3.089*
Event type	0.776	0.181	4.277**
Condition Non-verbal:Language	−0.530	0.283	−1.874 (p = .06)
Condition Verbal:Language	0.575	0.276	2.081*
Condition:Event type	−0.175	0.118	−1.490
Language:Event type	−0.169	0.093	−1.809 (p = .07)
Condition:Language:Event type	0.058	0.093	0.619

$^*p < .05$; $^{**}p < .001$.

speakers show a clear "verbal boost" when it comes to attention toward the Action depicted in the event stimuli. Overall, the condition effect (verbal boost on attention to the action) was driven by the Estonian group; Dutch participants displayed similar fixation behavior in the two conditions.

The exploratory analysis on Action AoI fixations in the Estonian verbal condition showed no significant effect of Case (β = −0.308, *SE* = 0.329, z = −0.936 ns). The use of the accusative specifically did not lead to enhanced action/object fixations.

5.6. Memory data

Fig. 6 below depicts accuracy scores for recognition memory in verbal and non-verbal encoding conditions.

Table 3 below shows the results of the mixed effect logistic regression model on memory accuracy.

The analysis showed a main effect of Condition, indicating overall higher memory performance after Verbal encoding. There was a significant Language by Condition interaction: In the Verbal experiment, Estonian participants outperformed the Dutch in memory of video endings, but in the Non-verbal experiment the pattern was reversed. However, Estonian participants thus displayed a verbal advantage for memory (superior performance after verbalization), and event memory was similar in the two conditions in Dutch participants. There were no effects of Event type.

The exploratory analysis on memory accuracy in relation to the use of specific case markers in the Estonian verbal condition showed no significant effect of the factor Case (β = 0.260, *SE* = 0.154, z = 1.685, ns).

6. Discussion

Our data show general cognitive principles and language-specific influences in how people perceived and memorized causative events involving agents inflicting varying

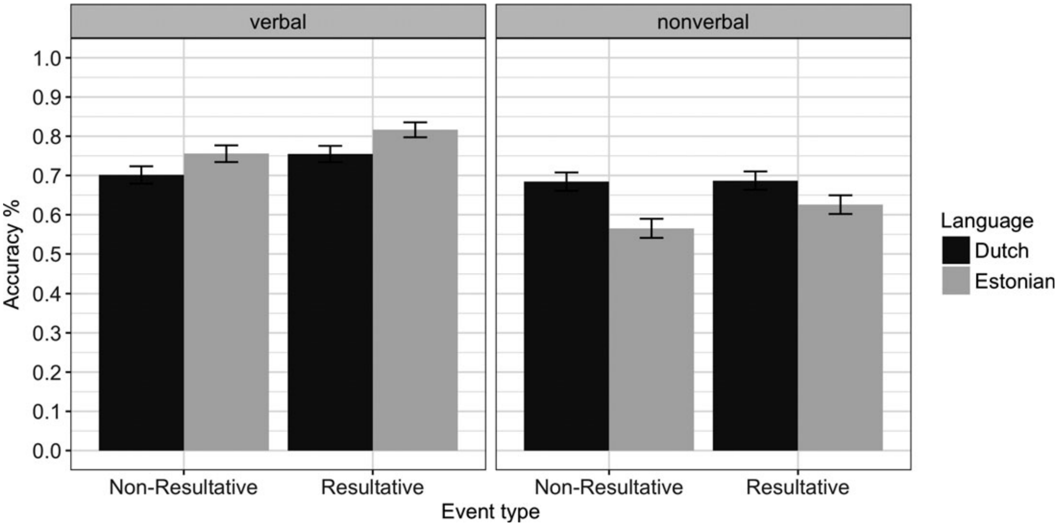


Fig. 6. Accuracy of recognition memory after verbal and non-verbal encoding, for each Language and Event type (error bars indicate ±*SE*).

Table 3
Output of the model on accuracy of event memory (glmer (accuracy ~ condition * language * event type – language + (1 + event type|participant) + (1 + condition|item), family = binomial)

Fixed Effect	*β* Estimate	*SE*	*z*-Value
Intercept	0.922	0.080	11.567**
Condition	0.314	0.058	5.39**
Event type	0.116	0.070	1.643
Condition Non-verbal:Language	−0.204	0.075	−2.717*
Condition Verbal:Language	0.169	0.078	2.148*
Condition:Event type	0.048	0.045	1.078
Language:Event type	0.042	0.040	1.057
Condition:Language:Event type	−0.020	0.040	−0.516

*p < .05; **p < .001.

degrees of causal change upon objects. Regardless of language background and demands of the task, participants allocated more attention toward the action and the object depicted in resultative, compared to non-resultative events, close to video offset. This region of interest and time window contain critical information on the event's result. Resultative events show a high degree of object affectedness, and the visual states of the objects in them imply that there is a high likelihood that the event will culminate with a full result, that is, a full change in state. Critically, whether or not this change in state would actually be achieved was visible only during the crucial analysis time window (at the end of the videos it would become clear whether the action would be ceased or remain ongoing, with consequences for the resultant state of the object). Non-resultative events involved objects which were not saliently changed during the time span depicted in the videos, regardless of ceased or ongoing action at video offset. Hence, the final unfolding of resultative event videos conveyed less predictable, and thus, more attentionally grabbing information compared to non-resultative event videos. Specifically, the precise way in which the action concluded had consequences for the qualitative properties of the objects involved, whereas it did not drastically affect object properties in non-resultative events. Arguably, thus, the difference between a ceased and ongoing resultative event is visually larger (i.e., it also involves differences in the properties of the object shown) than the same manipulation for a non-resultative event. Therefore, video endings of the former event type attracted more attention. In the subsequent recognition memory task, however, participants' memory of video endings was not significantly better for resultative events, suggesting that the visual features that attracted most attention online were not necessarily prioritized in participants' event memory and did not form a critical part of their eventual representation of the event.

We interpret these patterns as in line with Event segmentation theory (e.g., Shipley & Zacks, 2008), which proposes that moments and elements of accumulating change in an event should be highly salient to a perceiver. During the events' final unfolding, people were indeed more attentive when it was not entirely clear what would happen next, specifically, in relation to the object in the event; as mentioned above, the specific video ending affected the end state of the object in resultative, but not non-resultative events, while keeping the variability concerning the dynamics of the action (ceased or ongoing) constant across event types. We take these findings to show cognitive prominence of object states: Recent views put forward the idea that event end states and object affordances, specifically, degree of state change, are central dimensions in event processing and understanding (Altmann, 2017; Hindy, Altmann, Solomon, & Thompson-Schill, 2013; Solomon et al., 2015). This fits with the proposal that action concepts in the brain by default include information on the effect of the action on the patient involved. Word order preferences in the languages of the world, where the verb and the object are tightly connected given their inherent relatedness, can be traced back to this general principle of our neuro-cognitive architecture (Kemmerer, 2012). In this light, causative events are centered on the histories or "trajectories of change" that are carried by the objects in them. This puts a different weight on the relative importance of the elements of which an event is composed; other theories have, for example, proposed agents, as the actors causing the changes in an event, or the actions themselves, specifying the type of change, as fundamental primitives of event comprehension (e.g., Bickel, Witzlack-Makarevich, Choudhary, Schlesewsky, & Bornkessel-Schlesewsky, 2015; Kemmerer, 2012; Zwaan & Radvansky, 1998). Similar to the attentional pattern for state change in objects obtained here, in the domain of space, a universal bias toward goals in motion events has been proposed (a source-goal asymmetry, specifically). This has been shown in linguistic description, but also more globally, in, for example, perceptual change detection tasks (e.g., Lakusta & Landau, 2012; Papafragou, 2010; Regier & Zheng, 2007). Event types and dimensions with a high degree of depicted change seem to be universally salient visual features, regardless of the specific type of change involved (whether spatial or causal). Note, though, that the actual reaching of the event boundary (ceased action and a full result) was not the driving factor as our additional form-specific analysis suggests: The accusative itself does not explain the overall pattern. Rather, the unpredictability and uncertainty regarding partial/full event results seem to attract viewers' attention (in line with Event segmentation theory).

To specifically explore the importance of object states further, it would be interesting to tease apart the relative contribution of the dynamic action and the object itself (collapsed into one area of interest in the eye tracking analyses) to the attention patterns we obtained here. With the current stimulus set-up, this is unfortunately not possible, given a spatial overlap between the actor's hands engaged in the action and the affected object in all cases. We argue, however, that the differentiation between resultative and non-resultative events in the present design specifically isolates properties of the object: the ceased-ongoing manipulation implies object state change in the former, but not the latter event type. This allows us to infer that this event element is responsible for the increased attention to resultative events. Thus, extending the work on event comprehension on the basis of sentence materials (e.g., Hindy et al., 2012; Solomon et al., 2015), we argue for general cognitive saliency of the changes in state that an object undergoes in an event, during information uptake in language production (verbal scene encoding), and, importantly, also during visual event processing more broadly (non-verbal scene encoding).

However, in contrast to what existing theories might predict, we did not find resultative event endings to be prioritized in one's memory representation of the event. Note that, with the current manipulation of event type, we target representations that involve the *degree* of object state change, arguably tapping into a very subtle contrast (high vs. low saliency of change) which may not be substantiated in a memory representation given the degree of detail involved. Further research is required to shed light on how the various dimensions of an event and the distinctions within them (e.g., specific details of agents, actions, and patients; the dimensions of state change, causality, temporality, etc.) are represented in event memory.

Importantly, we employed a cross-linguistic comparison to shed light on how the grammatical marking of object affectedness influences perception and memory of causative events and their potential results. The comparison is based on the fact that, in contrast to Dutch, Estonian transitive event descriptions convey obligatory information on changes in state of objects (partial or full) through case markers on direct object nouns, hypothesized to lead to enhanced saliency of the results of events. During verbal encoding, Estonian participants allocated more attention to video endings than the Dutch, for *both* resultative and non-resultative event stimuli. Similarly, their memory of video endings was overall more accurate in the verbal condition. Dutch participants, on the other hand, did not display different behaviors in verbal and non-verbal conditions. These effects can be characterized as evidencing "thinking for speaking" processes in Estonian, that is, one's attention is driven toward characteristics of the visual input that are relevant for information retrieval during speech preparation. Interestingly, however, the Estonian verbal boost was found regardless of the specific saliency of changes in state in the events, as operationalized in our manipulation of event type. In addition, the exploratory form-specific analysis did not show an enhancing effect of accusative case marking on attention and memory. Both findings underline that the *specific* case marker retrieved by the Estonian speakers in the experiment made no significant difference: The fact that, in contrast to Dutch, Estonian speakers need to make explicit a perspective on perceived degree of object state change and the result of an event — no matter which one — triggered enhanced attention. Similarly, their event memory was boosted after verbalization of the scenes, independent of the specific forms uttered. This is a novel finding in the "thinking for speaking" literature: Whereas there is ample evidence for online attention biases induced by selection and retrieval of *specific* linguistic forms and structures (e.g., Bunger et al., 2016; Flecken, Carroll, Weimar, & van Stutterheim, 2015; Hendriks, Hickmann, & Demagny, 2008; Papafragou & Selimis, 2010; Papafragou et al., 2008; Slobin, 2006; Soroli & Hickmann, 2010), the present data suggest that these effects are not limited to such contexts. Here, a language-specific effect surfaces regardless of the specific form in the linguistic output; it is driven by general requirements of Estonian grammar, namely, the fact that a speaker has to specify a(ny) perspective on an event's result when processing causative events for verbalization.

In non-verbal event memory, we find an unexpected reversed language effect (and a similar numerical trend in the eye movement data): Event memory in Estonian was clearly boosted under verbal task demands, but after non-verbal encoding their performance on the event recognition task was lower than in Dutch participants. A speculative explanation for this pattern is that Estonian speakers may generally have more of an opportunity to rely on linguistic resources for committing events to memory. The Estonian linguistic system is morphologically rich compared to Dutch; for example, it has an elaborate case marking system, encoding event-related distinctions in terms of spatial relations, agent-patient relations, and transitivity, potentially providing speakers with a systematic scaffold for structuring event representations. This might be a very efficient tool to aid event memory. Then, in cases where the system is not readily available (as in the present non-verbal task), participants could not rely on their default strategy, leading to impaired performance. Alternatively, one might speculate that, despite similar visual-spatial and verbal working memory, and high performance on the distracter task, the non-verbal dual-task may have affected attentional capacities in the two groups differently. This would apply specifically to attentional resources left for the inspection of video *endings*, given that the task-relevant auditory cues were always played during the mid-phases of the (filler) video clips. Both hypotheses warrant further exploration in future studies.

Overall, our findings do not evidence linguistic relativity effects: The Estonian attentional boost in relation to event results is restricted to verbalization contexts exclusively. It does not seem to be the case that grammatical concepts specifically, given habitual (obligatory) activation in given contexts, will lead to strong and global effects on cognitive processing by default (e.g., Lucy, 1992). This is different from relativity effects found in relation to grammatical aspect and motion event cognition, where global, linguistically driven cognitive biases have been observed (e.g., Athanasopoulos & Bylund, 2013; Flecken, Athanasopoulos, et al., 2015). There are two potential explanations for these different findings: First, looking at the descriptions, although overall the experimental manipulations triggered the expected Estonian case alternation ensuring the validity of our central hypothesis empirically, there is variability in case marking choices. As mentioned above, several factors could have influenced speakers' choices for the one over the other case marker, leading to variability within the language system of interest and thus providing us with no black and white cross-linguistic contrast as test case for language effects on cognition (see Montero-Melis et al., 2017). However, there is a second plausible explanation: Note that the specific carriers of perspectives on event results in Estonian are nominal in nature; that is, partitive or accusative case is specified on object *nouns*. This is different from grammatical aspect, which also specifies (temporal) event perspectives, but which is marked verbally. The verbal nature of aspect entails that this type of perspective-taking takes place during *early* processing stages in sentence production, that is, information selection during event conceptualization (Levelt, 1989; see von Stutterheim & Nüse, 2003). Given SVO word order and because of the incremental nature of language production, linguistic retrieval of the object noun specifically takes place later during the time course of visual scene processing. The specification of verbally marked aspectual event perspectives may thus drive highly automatized and deeply entrenched cognitive effects that are more likely to surface independent of overt verbalization instructions, compared to the present linguistic phenomenon. To shed light on this issue, it would be highly interesting to experimentally explore potential biases toward event results in languages that mark event results verbally through, for example, an aspectual opposition between imperfective and perfective verbal morphology (see, e.g., Dahl, 2000). Perfective aspect in particular encodes event completion and highlights event end states, potentially leading to prominence in attentional processing of the relevant dimension, which in turn could lead to stronger general, that is, task-independent, cognitive saliency of event results and object end states. Ünal, Pinto, Bunger, and Papafragou (2016) investigate event memory in relation to verbal marking of evidentiality in Turkish, and the absence thereof in English. This study, however, does not report language-specific influences on memory. In order to shed light on the role of verbal versus nominal marking of event perspectives on attention and memory, a study comparing the two for the same event dimension is warranted.

Finally, it is noteworthy that, unlike many previous studies, we did not adopt a verbal interference manipulation during non-verbal encoding. Verbal interference is a procedure commonly used to prevent the use of covert verbal strategies (e.g., Trueswell & Papafragou, 2010). It is, however, unclear what phases in the language processing system are affected by this manipulation: The phonological loop is occupied, but what about other formulation processes and conceptualization? Also, it is difficult to control for the degree of complexity and cognitive load added to the main task (Perry & Lupyan, 2013), and, critically, it does not ensure participants' attention to stimulus contents. Importantly, the linguistic distinction of interest here concerns a subtle element of an event description; it is unlikely that such linguistic structures are part of a potential covert scene verbalization strategy. The present auditory distracter task did modulate attention and memory as compared to the verbal condition, showing that it at least attenuated such potential strategies, while ensuring attention paid to the contents of the videos that people were watching.

7. Conclusions

Overall, we demonstrate a general cognitive principle of enhanced attention to the end states of actions and objects in resultative, compared to non-resultative events, in viewers with different language backgrounds. In our design, we manipulated the degree of causal change inflicted upon an object by an agent in an event, so as to show events that either involved visually salient changes in state of objects and had the potential of concluding with a full result (resultative events) or those that ended with a partial result only. Our data also evidence a language-specific verbal boost on the saliency of event results, both during online event encoding and in offline recognition memory. The language effects were driven by the grammatical requirements of Estonian, a language that marks object affectedness and, with that, the perceiver's viewpoint on the result of an event, through case morphology on direct object nouns.

In all, this specific experimental design tapping into two aspects of event cognition (attention during and memory after event encoding) under varied task demands, and across different languages, is a highly useful tool for advancing our understanding of human event cognition.

Acknowledgments

We thank three anonymous reviewers for their constructive feedback on this manuscript. In addition, we thank Ercenur Unal for commenting on an earlier draft of this paper.

Notes

1. Dutch does make use of resultative verb particles and prepositional phrases to mark event results (e.g., "the man ate the apple up"; van Hout 1998). We report a linguistic analysis of the use of such means in the event description data of the two groups, shedding light on potential language differences. In addition, for Dutch we coded the use of tense-aspect contrasts which may provide information on event results (present perfect "he has eaten the apple" vs. progressive aspect "he was eating the apple").
2. We aimed to include a sample of 24 participants (six in each of four pseudo-randomized lists) in each encoding condition, for each language group, based on common practices in cross-linguistic studies of event processing. When technical

issues were noticed during the experimental session, participants were replaced with another person. Some technical, performance- or background-related issues were only discovered after data transcription and preprocessing. It was decided to not test additional participants at this stage anymore, ensuring that data collection was not influenced by data inspection or analysis.

3. Data from five participants were discarded due to high tracking loss (tracking ratio <70%). Five further participants were excluded based on incomplete event descriptions (producing nominalizations of the action, e.g., *the peeling of a banana*), thus not fulfilling the task of producing full-fledged sentences).
4. One participant was excluded due to low tracking ratio (<70%); another one was excluded because (s)he turned out to be Russian-Estonian early bilingual.
5. Participants were excluded due to high tracking loss (ratio <70%).
6. Data from two Dutch speakers were excluded due to a technical problem during the memory task.
7. Stimuli are available publicly at https://osf.io/uyxtg/
8. Events involved either partial object state change (e.g., grate cheese, knit scarf; grating some cheese from a large chunk, knitting a scarf with only a small part of the scarf done), or only superficial, and thus not visually salient change (whisk cream, polish glasses — the effect of the action is not visible on the object), or no change at all (read a book, measure a box). The objects' states give away that it is not likely that the event will produce a full result, in the form of a fully changed object; as such, they are not expected to elicit accusative case-marking in Estonian.
9. The Dutch *aan het* construction (*een vrouw is een aardappel aan het schillen* "a woman is peeling a potato") marks progressive aspect, specifying an event as currently "ongoing" (as opposed to having reached a state of completion). The construction is, however, not fully grammaticalized and use is optional to describe ongoing actions (Behrens, Flecken, & Carroll, 2013).
10. The Agent AoI was mainly matched with the face of the actor, as facial features were most relevant for the identification of the agent in our stimuli. The Action AoI included both the ongoing action of the hands (and potential instrument) plus the affected object, as the present set-up does not allow to disentangle looks to these elements (hands plus instrument are never spatially distant from object).
11. We reasoned, given that video clips ended and disappeared from the screen after 3,000 ms, participants would not move their eyes from the region on the screen they were focusing on in that moment for at least an additional 200 ms (as it takes about that time to plan and launch a saccade). Inspecting the plots in Appendix B, indeed, looks to both AoIs drop after video offset, with fixations outside both AoIs exceeding looks to any predefined AoI from 300 ms postvideo offset. Hence, we analyzed looks from 300 ms before video offset (point in time during which culmination became visible) until 300 ms after video offset.
12. Data from one Dutch participant in the verbal condition were excluded due to a technical problem (failure to save the output file).

13. This category subsumes instances of intransitive sentences (without any object) and sentences with indirect objects (marked with a different case).

References

Ackerman, F., & Moore, J. (2001). *Proto-properties and grammatical encoding: A correspondence theory of argument selection*. Stanford, CA: CSLI.

Altmann, G. (2017). Abstraction and generalization in statistical learning: Implications for the relationship between semantic types and episodic tokens. *Philosophical Transactions of the Royal Society B*, *372*, 20160060.

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

Athanasapoulos, 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., & 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.

Behrens, B., Flecken, M., & Carroll, M. (2013). Progressive attraction: On the use and grammaticalization of progressive aspect in Dutch, Norwegian and German. *Journal of Germanic Linguistics*, *25*(2), 95–136.

Bickel, B., Witzlack-Makarevich, A., Choudhary, K., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLoS ONE*, *10*(8), e0132819.

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.

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition*, *118*(1), 123–129.

Bunger, A., Skordos, D., Trueswell, J., & Papafragou, A. (2016). How children and adults encode causative events cross-linguistically: Implications for language production and attention. *Language, Cognition and Neuroscience*, *31*(8), 1015–1037.

Carroll, M., Weimar, K., Flecken, M., Lambert, M., & von Stutterheim, C. (2012). Tracing trajectories: Motion event construal by advanced L2 French-English and L2 French-German speakers. *Language, Interaction and Acquisition*, *3*(2), 202–230.

Casasanto, D. (2008). Who's afraid of the Big Bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning*, *58*, 63–79.

Dahl, O. (2000). *Tense and aspect in the languages of Europe*. Berlin: de Gruyter.

Fausey, C., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, *18*(1), 150–157.

Filipović, L. (2011). Speaking and remembering in one or two languages: Bilingual vs. monolingual lexicalization and memory for motion events. *International Journal of Bilingualism*, *15*(4), 466–485.

Finkbeiner, M., Nicol, J., Greth, D., & Nakamura, K. (2002). The role of language in memory for actions. *Journal of Psycholinguistic Research*, *31*(5), 447–457.

Flecken, M., Athanasopoulos, P., Kuipers, J. R., & Thierry, G. (2015a). On the road to somewhere: Brain potentials reflect language effects on motion event perception. *Cognition*, *141*, 41–51.

Flecken, M., Carroll, M., Weimar, K., & von Stutterheim, C. (2015b). Driving along the road or heading for the village? Conceptual differences underlying motion event encoding in French, German, and French-German L2 users. *Modern Language Journal*, *99*, 100–122.

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.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, *83*(1), 49–79.

Gumperz, J., & Levinson, S. (Eds.) (1996). *Rethinking linguistic relativity*. Cambridge, UK: Cambridge University Press.

Han, Z-H., & Cadierno, T. (2010). *Linguistic relativity in SLA*. Clevedon, UK: Multilingual Matters.

Hendriks, H., Hickmann, M., & Demagny, A.-C. (2008). How adult English learners of French express caused motion: A comparison with English and French natives. *Acquisition et Interaction en Langue Étrangère*, *27*, 15–41.

Hindy, N., Altmann, G., Kalenik, E., & Thompson-Schill, S. (2012). The effect of object state-changes on event processing: Do objects compete with themselves? *Journal of Neuroscience*, *32*(17), 5795–5803.

Hindy, N., Solomon, S., Altmann, G., & Thompson-Schill, S. (2013). A cortical network for the encoding of object change. *Cerebral Cortex*, *25*, 884–894.

Imai, M., Schalk, L., Saalbach, H., & Okada, H. (2014). All giraffes have female- specific properties: Influence of grammatical gender on deductive reasoning about sex-specific properties in German speakers. *Cognitive Science*, *38*(3), 514–536.

Ji, Y., Hendriks, H., & Hickmann, M. (2011). How children express caused motion events in Chinese and English: Universal and language-specific influences. *Lingua*, *121*(12), 1796–1819.

Kemmerer, D. (2012). The cross-linguistic prevalence of SOV and SVO word orders reflects the sequential and hierarchical representation of action in Broca's area. *Language and Linguistics Compass*, *6*, 50–66.

Kersten, A. W., Meissner, C. A., Lechuga, J., Schwartz, B. L., Albrechtsen, J. S., & Iglesias, A. (2010). English speakers attend more strongly than Spanish speakers to manner of motion when classifying novel objects and events. *Journal of Experimental Psychology: General*, *139*(4), 638–653.

Kessels, R. P., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., & De Haan, E. H. (2000). The Corsi block-tapping task: Standardization and normative data. *Applied Neuropsychology*, *7*(4), 252–258.

Kiparsky, P. (1998). Partitive case and aspect. In M. Butt & W. Geuder (Eds.), *The projection of arguments: Lexical and compositional factors* (pp. 265–308). Stanford, CA: CSLI.

Lakusta, L., & Landau, B. (2012). Language and memory for motion events: Origins of the asymmetry between source and goal paths. *Cognitive Science*, *36*(3), 517–544.

Lees, A. (2004). Partitive-accusative alternations in Balto-Finnic languages. In C. Muskovsky (Ed.), *Proceedings of the 2003 Conference of the Australian Linguistic Society*. Available at http://www.als.asn.au

Levelt, W. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Lucy, J. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge, UK: Cambridge University Press.

Montero-Melis, G., Eisenbeiss, S., Narasimhan, B., Ibarretxe-Antuñano, I., Kita, S., Kopecka, A., & Bohnemeyer, J. (2017). Satellite- vs verb-framing underpredicts nonverbal motion categorization: Insights from a large language sample and simulations. *Cognitive Semantics*, *3*(1), 36–61.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28.

Papafragou, A. (2010). Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive Science*, *34*, 1064–1092.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements *Cognition*, *108*(1), 155–184.

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, *84*(2), 189–219.

Papafragou, A., & Selimis, S. (2010). Event categorisation and language: A cross-linguistic study of motion. *Language and Cognitive Processes*, *25*(2), 224–260.

Pavlenko, A., & Volynsky, M. (2015). Motion encoding in Russian and English: Moving beyond Talmy's typology. *The Modern Language Journal*, *99*(S1), 32–48.

Perry, L. K., & Lupyan, G. (2013). What the online manipulation of linguistic activity can tell us about language and thought. *Frontiers in Behavioral Neuroscience*. https://doi.org/10.3389/fnbeh.2013.00122

Radvansky, G., & Zacks, J. (2011). Event perception. *WIREs Cognitive Science*, *2*(6), 608–620.

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

Regier, T., & Zheng, M. (2007). Attention to endpoints: A cross-linguistic constraint on spatial meaning. *Cognitive Science*, *31*, 705–719.

Shipley, T., & Zacks, J. (2008). *Understanding events: From perception to action*. Oxford, UK: Oxford University Press.

Slobin, D. (1996). From thought and language to thinking for speaking. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, UK: Cambridge University Press.

Slobin, D. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind. Advances in the study of language and thought* (pp. 157–191). Cambridge, MA: MIT Press.

Slobin, D. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 59–81). Amsterdam: John Benjamins.

Solomon, S., Hindy, N., Altmann, G., & Thompson-Schill, S. (2015). Competition between mutually exclusive object states in event comprehension. *Journal of Cognitive Neuroscience*, *27*(12), 2324–2338.

Soroli, E., & Hickmann, M. (2010). Language and spatial representations in French and in English: Evidence from eye-movements. In G. Marotta, A. Lenci, L. Meini, & F. Rovai (Eds.), *Space in language* (pp. 581–597). Pisa, Italy: Editrice Testi Scientifici.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description. Grammatical categories and the lexicon* (pp. 57–149). Cambridge, UK: Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

Tamm, A. (2004). Estonian transitive verb classes, object case, and progressive. *Nordlyd*, *31*(4), 639–653.

Tamm, A. (2007). Perfectivity, telicity and Estonian verbs. *Nordic Journal of Linguistics*, *30*(2), 229–255.

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.

van Hout, A. (1998). *Event semantics of verb frame alternations. A case study of Dutch and its acquisition*. New York: Routledge.

von Stutterheim, C., & Nüse, R. (2003). Processes of conceptualization in language production: Language-specific perspectives and event construal. *Linguistics*, *41*(5), 831–881.

Ünal, E., Pinto, A., Bunger, A., & Papafragou, A. (2016). Monitoring sources of event memories: A cross-linguistic investigation. *Journal of Memory and Language*, *87*, 157–176.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253–265.

Wolff, P., & Ventura, T. (2009). When Russians learn English: How the semantics of causation may change. *Bilingualism: Language and Cognition*, *12*(2), 153–176.

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., Hink, R. F., & Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology*, *33*(1), 101–111.

Zacks, J., Braver, T., Sheridan, M., Donaldson, D., Snyder, A., Ollinger, J., Buckner, R., & Raichle, M. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*, 651–655.

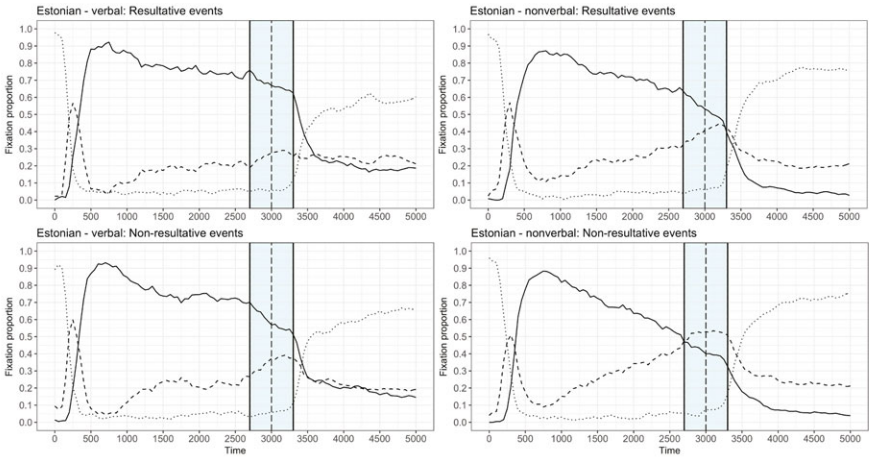
Zacks, J. M., & Swallow, K. (2007). Event segmentation. *Current Directions in Psychological Science*, *16*(2), 80–84.

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

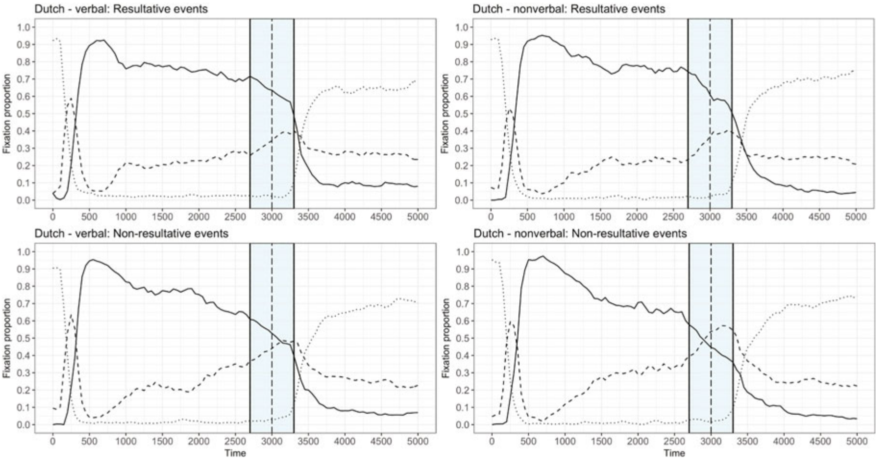
Appendix A: List of stimulus items

Resultative Events (N = 18)	Non-resultative Events (N = 18)	Fillers (N = 18)
Break chocolate into pieces	Whisk cream	Yawn
Build a Lego tower	Measure a box	Stretch arms
Crunch a sheet of paper	Shuffle cards	Sleep with head on table
Cut an apple in half	Play a drum	Juggle
Cut a circle out of paper	Read a book	Look in a mirror
Draw a flower	Stir soup	Blow one’s nose
Fold a paper airplane	Salt a soup	Use a calculator
Cut paper in half	Polish glasses	Talk on the phone
Open a can	Clean a mirror	Throw paper in a bin
Open a jar	Polish a glass	Bandage one’s hand
Open a letter	Wipe a table	Dance
Peel a banana	Rub a knife with a cloth	Give a flower
Peel a mandarin	Cut fingernails	Put a book on the table
Peel a potato	Knit a scarf	Put a book on one’s head
Pour a glass of coke	Pour water from a flask	Put on a hat
Make a jigsaw puzzle	Grate cheese	Shake hands
Roll wool into a ball	Staple papers	Take someone’s glasses
Tear paper in half	Rub lotion on hands	Throw a ball

Appendix B: Fixation proportions in Agent and Action areas of interest, and outside of both, in Estonian participants



Fixation proportions in Agent and Action areas of interest, and outside of both, in Dutch participants



Solid lines: Action AoI; dashed lines: Agent AoI; dotted lines: Outside of both AoIs.

X-axis shows time from stimulus onset until 5,000 ms, in 50 ms bins (dashed vertical line at 3,000 ms is stimulus offset; shaded area represents time window for analysis).