Journal of Memory and Language 149 (2026) 104752

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Remembering times ahead: The effect of linguistic framing on representational momentum in state-change events

Xueyi Yao [a], Natalia Jardón [a,b], Jonathan Kominsky [a], Eva Wittenberg [a,*]

[a] Department of Cognitive Science, Central European University, Quellenstraße 51, 1100 Wien, Austria
[b] Departament de Filologia Catalana, Edifici B, Despatx B9/0050, Facultat de Filosofia i Lletres, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

ARTICLE INFO

Keywords:
Grammatical aspect
Event memory
Representational momentum
State-change events

ABSTRACT

When people remember scenes, they can draw from different sources to reconstruct their memory, such as what they saw, and linguistic descriptions of what they saw. How human memory for events is impacted by linguistic encoding has been a long-standing question in cognitive science. Here, we ask whether language permeates memory of state-change events. We use the visual phenomenon of Representational Momentum, in which people misremember an event they saw (e.g., a melting ice cube) as more completed than it actually was. After three successful replications of the Representational Momentum effect for state changes (total N = 150), in two further experiments, participants (total N = 400) watched partial videos of state change events that were followed by linguistic descriptions in either imperfective aspect ("the ice was melting") or perfective aspect ("the ice has melted"). As predicted by linguistic theories, perfective aspect increased the Representational Momentum effect, compared to imperfective aspect. In addition, linguistic encoding improved overall memory accuracy. Our work presents the first evidence that the Representational Momentum effect for state change events is sensitive to linguistic framing, specifically the use of grammatical aspect. More generally, our data contribute to a model of cognition in which different kinds of representations interact in memory.

Introduction

The world we live in unfolds through dynamic events: As you look out the window, you might see a teenager kicking a ball, or a chunk of snow that is melting under the midday sun. Integral to these everyday experiences is the notion of change: change of location of an object relative to its external environment (e.g., the ball moved by the kick), as well as change of state of an object's internal properties (e.g., the snow becoming liquified). Understanding how we perceive and interpret these changes offers valuable insight into the mechanisms of visual cognition.

One foundational finding in higher-level visual cognition is the so-called representational momentum effect (see Hubbard, 2005, for a review). Representational momentum refers to the phenomenon of people extrapolating the positions of moving objects in memory tasks: People misremember objects as being displaced forward along their trajectories. Recently, this phenomenon has also been found in state-change events (Hafri et al., 2022), for which individuals remember the objects' states to be more changed than they actually were.

Importantly, a long tradition of research has argued that visual processes are, at one point, also susceptible to top-down conceptual information, such as linguistic framing (e.g., Forder & Lupyan, 2019; Lupyan et al., 2020; Maier & Abdel Rahman, 2019; Reed & Vinson, 1996). While psycholinguistic research has shown that language may influence how events are conceptualized and remembered, it remains an open question whether and how a reliable and automatic visual effect, such as representational momentum, manifests under conditions of linguistic framing. In this study, we ask whether grammatical aspect, which has been independently found to affect how people conceptualize dynamic events (Anderson et al., 2013; Madden & Zwaan, 2003; Minor et al., 2022; Misersky et al., 2021; Zhou et al., 2014), could permeate the fundamental cognitive processes of recalling mentally extrapolated state-change events.

The phenomenon of representational momentum

When a ball is kicked, it changes its location. To catch it, we must not only track its current position but also predict where it will go next (Wolpert & Flanagan, 2001; Witney et al., 2001). A key finding in visual cognition reveals that this 'forward momentum' is encoded in object representation, leading to memory distortions: People tend to

* Corresponding author at: Department of Cognitive Science, Central European University, Quellenstraße 51, 1100 Wien, Austria.
E-mail address: wittenberge@ceu.edu (E. Wittenberg).

https://doi.org/10.1016/j.jml.2026.104752
Received 14 May 2025; Received in revised form 26 February 2026; Accepted 27 February 2026
Available online 20 March 2026

misremember objects as being slightly further along in their trajectories, a phenomenon known as 'representational momentum' (Freyd, 1983; Freyd & Finke, 1984; Hubbard, 2005).

Freyd and Finke (1984) first demonstrated representational momentum using a rotating rectangle paradigm (Fig. 1). Observers viewed three sequentially presented rectangles (inducing stimuli), followed by a test stimulus after a brief retention interval. When the sequence of inducing stimuli implied rotation in a consistent direction with a consistent speed, observers were more likely to judge the test stimulus as 'same' when it was slightly further along that implied rotation, indicating a memory distortion in the direction of motion. This effect has since been replicated across various paradigms, including some involving multiple targets (Finke & Freyd, 1985).

The automatic nature of representational momentum has led some researchers to propose that it is a low-level visual illusion arising from the properties of the visual system (e.g., Bertamini, 2002; Kerzel, 2000; Kerzel et al., 2001). Kerzel (2000) found that when tracking a moving object, participants' eyes tend to overshoot its final position along the trajectory of the object. Similarly, Müsseler et al. (1999) reported that memory for an object's location is biased toward the fovea, the central retinal region responsible for sharp vision. These combined effects of pursuit eye movements and foveal bias led Kerzel (2000; Kerzel et al., 2001) to propose that forward displacement results from oculomotor behavior.

However, others argue that representational momentum is better understood as a higher-level cognitive phenomenon. Finke and Freyd (1985) and Finke et al. (1986) proposed the 'momentum' metaphor, drawing an analogy to physical momentum: In biological motion, moving objects do not stop instantly. Mental representations take this into account and extend beyond an object's final position, causing an effect of forward displacement that supports anticipation and motor regulation. Freyd (1987, 1992, 1993) further argued that representational momentum reflects the drive to spatiotemporal coherence—the alignment between the external physical world and internal mental representations. To achieve coherence, mental representations need to be dynamic, inherently integrating the passage of time to mirror real-world motion, thereby producing forward displacement.

Researchers have previously speculated that representational momentum shouldn't be restricted to movement, and the mind might extrapolate any transformation forward in time (Finke et al., 1986). Recent work has confirmed this speculation and found the representational momentum effect in state-change events as well. In Hafri et al. (2022), participants viewed videos of objects undergoing state changes (e.g., ice melting, or a log burning) and were asked to choose the last frame they saw from each video by pulling a slider. The results showed that participants consistently selected frames that depicted the objects as more transformed than they actually were. This forward bias persisted even when state-change animations played in reverse, meaning that participants remembered the object as more unchanged than it really was.

Hafri et al.'s (2022) work demonstrates that our mind dynamically represents physical changes of objects and incorporates their probable future states into memory. It supports the mental representation account (Freyd, 1987, 1992, 1993) and raises the possibility that the representational momentum effect might not be fully encapsulated if it is based not on low-level perceptual processing of motion, but on more abstract representations such as change of state.

Indeed, there is evidence showing that conceptual knowledge about the identity of a target influences forward displacement in motion events (Reed & Vinson, 1996; Vinson & Reed, 2002). For instance, an upward-moving stimulus labeled "rocket" exhibits greater forward displacement than an identical stimulus labeled "cathedral". This work supports the argument that representational momentum in motion events reflects a cognitive process in which individuals integrate knowledge, expectations and beliefs to mentally extrapolate the position of a moving object, leading to a post-perceptual false memory. However, thus far the only demonstration of this cognitive influence has focused on the properties of objects, rather than features of the event representation as a whole.

The role of language in event conceptualization and event memory

Language affects the way events are conceptualized (e.g., Fausey & Boroditsky, 2011; Marx & Wittenberg, 2025; Marx et al., 2025; Wittenberg & Levy, 2017) and remembered (e.g., Feist & Gentner, 2007;

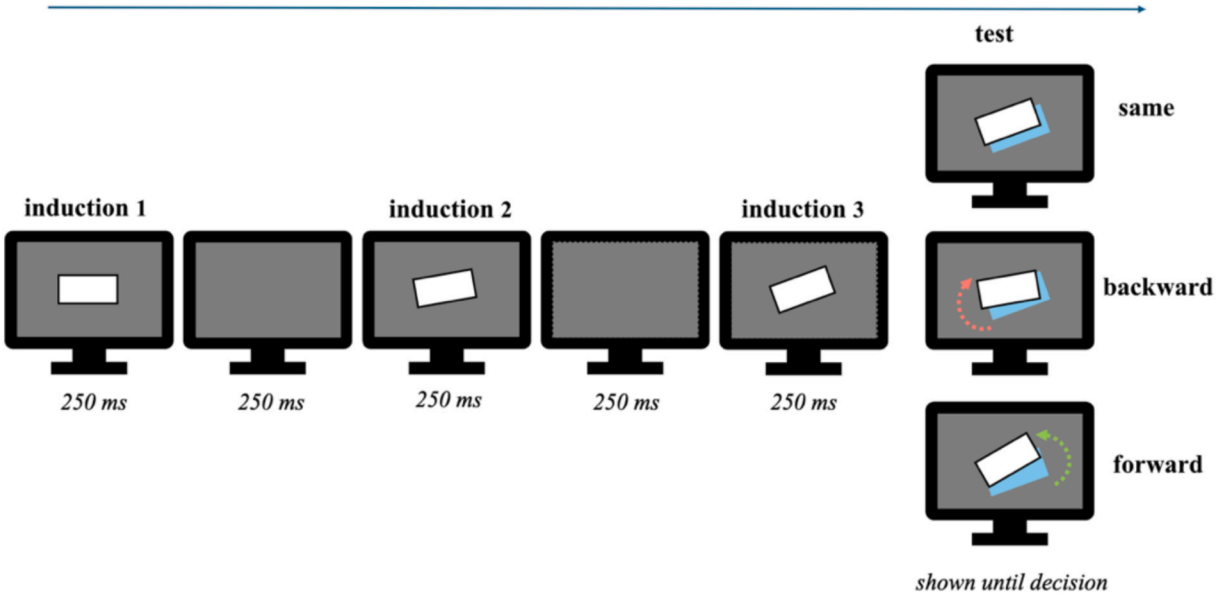

Fig. 1. Illustration of display sequences used in Freyd & Finke (1984), shown in a counterclockwise rotation. Each trial began with three induction displays, each shown for 250 ms, with 250 ms between them. These displays implied a consistent counterclockwise rotation. In the test phase, participants viewed one of three possible test displays and decided based on them. In the forward condition, the test stimulus (white) continued the implied rotation. In the backward condition, the test stimulus reversed direction. In the same condition, the test stimulus matched the third induction stimulus. The test display remained onscreen until participants responded. Note: The blue rectangle was used to indicate the last position of the true test stimulus and was not shown during actual test trials; it is shown here for demonstration purposes only. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Gentner & Loftus, 1979; Loftus & Palmer, 1974; Marian & Neisser, 2000; Santin et al., 2021; Wang & Gennari, 2019; but see e.g., Gennari et al., 2002; Papafragou et al., 2008: Trueswell & Papafragou, 2010). Of particular interest in understanding how linguistic encoding affects event processing, construal, and memory, is grammatical aspect.

Grammatical aspect has been claimed to provide an internal temporal perspective on an event (Comrie, 1976; Smith, 1991; van Hout, 2016). Specifically, while perfective aspect presents the event as a complete whole, imperfective aspect presents the event as ongoing without specifying its limits. The prototypical way to express perfective aspect in English is through the simple past ('I danced'), while the prototypical way to express imperfective aspect in English is through the progressive ('I was dancing').

Previous experimental studies have confirmed the theoretical claims on the function of grammatical aspect in a number of paradigms and languages (Anderson et al., 2013; Becker et al., 2013; Bergen & Wheeler, 2009; Carreiras et al., 1997; Ferretti et al., 2007; Huette et al., 2014; Madden & Zwaan, 2003; Morrow, 1985; Morrow, 1990; Minor et al., 2022; Misersky et al., 2021; Magliano & Schleich, 2000; Truitt & Zwaan, 1997; von Stutterheim et al., 2012; Mozuraitis et al., 2013; Zhou et al., 2014). Specifically, imperfective aspect has been shown to increase access to event-internal information, such as locations (Ferretti et al., 2007), participants (Carreiras et al., 1997; Morrow, 1985; Morrow, 1990; Madden & Therriault, 2009) and instruments (Truitt & Zwaan, 1997). In other words, imperfective aspect provides a "lens" into the details of the internal stages of the event, leading to a higher availability of content associated with the event. Additionally, using imperfective aspect to indicate that an event is ongoing enhances the accessibility of that event during later stages of narrative processing (Magliano & Schleich, 2000), with this effect being particularly pronounced for accomplishment predicates such as "to build a castle" or "to listen to a song" (Becker et al., 2013).

In contrast, perfective aspect shifts people's focus onto the entire concluded event, as predicted from its semantics (Athanasopoulos & Bylund, 2013; Jardón et al., 2025; Madden & Zwaan, 2003; Morrow, 1985; Morrow, 1990; Misersky et al., 2021). In a foundational paper, Madden and Zwaan (2003) performed a series of experiments in which participants were exposed to sentences and pictures of different kinds of dynamic events, to test how grammatical aspect impacts the mental representation of events in English. Madden & Zwaan (2003) found that perfective sentences (e.g., "The man crossed the street") led to a preference for completed-event pictures, whereas imperfective sentences (e.g., "The man was crossing the street") did not influence preferences toward unfinished events. Follow-up reaction-time experiments revealed that perfective sentences prompted quicker responses to completed-event pictures, while imperfective sentences showed no such effect. These results suggest that the perfective aspect constrains the comprehenders' representation of the event as completed.

However, the completion of many events does not involve changes of location, like crossing the road, but changing the state of an object: For instance, in the melting of an ice cube, we look at the state of the ice cube, and whether it is completely changed into water. Tracking objects' states is central to linguistic and non-linguistic event comprehension and encoding (Altmann & Ekves, 2019; Hindy et al., 2012; Kang et al., 2020; Prystauka et al., 2024; Solomon et al., 2015), inasmuch as objects' initial, intermediate, and end states are activated during processing and in memory.

There is some initial evidence that linguistic framing, highlighting different states of an object, influences how events are remembered: Sakarias & Flecken (2019) tested Dutch and Estonian native speakers on their memory for state change events and found that Estonian case-marking, which grammatically encodes degrees of change (partial or fully change) in the object, improved event memory for recognizing end states compared to Dutch, which lacks this grammatical feature.

In the context of state change events, grammatical aspect has been used to linguistically access tracking an object's history, since aspect influences how salient an object's state is in our mind. For instance, Misersky et al. (2021) used a sentence-picture verification task and measured ERPs to test whether grammatical aspect can cue different phases of a state-change event. Participants read sentences in either perfective ("chopped the onion") or imperfective aspect ("was chopping the onion"), followed by images of objects in changed/resultant states (e.g., a chopped onion), unchanged states (e.g., a whole onion), or unrelated states (e.g., a cactus). The study found that images of objects in a changed state elicited a higher amplitude P300 component after sentences in perfective aspect compared to progressive aspect. These results suggest that perfective aspect cues a holistic event representation that includes the resultant state of the object, thereby construing the visual representation of the object's result state.

Misersky et al.'s (2021) study, however, did not include pictures corresponding to intermediate stages of change (e.g., an onion half chopped) and it consequently cannot inform us about the accessibility of the imperfective, and how it may direct participants' attention to objects undergoing a change. Thus, while we know from previous studies on state-change that linguistic framing (case-marking on the object, or grammatical aspect on the verb) can improve memory representations of end states, the association of a linguistic description with internal phases of change is less conclusive, beyond the general observation that imperfective aspect can draw attention towards events as they are unfolding (Carreiras et al., 1997; Ferretti et al., 2007; Truitt & Zwaan, 1997; Minor et al. 2022).

The present studies and predictions

In the present studies, we ask whether linguistic framing impacts representational momentum in state-change events. Previous work has found the effect of representational momentum in objects undergoing motion change and state change (Freyd & Finke, 1984; Hafri et al., 2022). Even though the nature of representational momentum in motion events is still under debate, that is, whether it originates in the encoding or in the retrieval stage, some studies have demonstrated that it seems not to be an encapsulated process. Rather, it is permeable by conceptual knowledge induced by linguistic labeling (Reed & Vinson, 1996; Vinson & Reed, 2002). This conceptual knowledge, however, relied on participants' physical intuitions about certain kinds of objects, instead of manipulating top-down information about the event itself.

Grammatical aspect, in contrast, maps onto an event structure, while concerning the representation of the object itself only as a consequence of the event's internal progression. Previous experimental studies have found that perfective aspect triggers a holistic representation of the event as a completed whole. In contrast, imperfective aspect foregrounds the in-progress stage of an event, thereby enhancing the accessibility of information relevant to the events. Specifically, studies that found evidence of imperfective aspect on object information (e.g., Misersky et al., 2021; Madden & Therriault, 2009) typically employed images depicting objects in either its initial or final state (e.g., an open vs. closed umbrella). However, such binary static depictions do not allow for a nuanced examination of how imperfective aspect may guide attention toward the internal progression of an event – especially in the case of state-change events, where objects undergo gradual transformation. We propose that using a continuous dynamic animation depicting state changes over time would offer clearer insight into how imperfective aspect modulates attentional focus on ongoing change. Simultaneously, this approach could help determine whether perfective aspect shapes event conceptualization as moving forward in time, without necessarily invoking categorical representation, a limitation inherent in previous methodologies relying on static binary images.

Our study aims to understand whether grammatical aspect could permeate the fundamental cognitive processes of mentally recalling extrapolated state-change events. Following independent findings on the effect of grammatical aspect on event conceptualization (Madden & Zwaan 2003; Madden & Therriault, 2009; Misersky et al. 2021), which

only used categorical completion in still images, we propose that grammatical aspect may also affect people's mental representations forward in time in a continuous manner. We reasoned that grammatical aspect will affect the process of mental extrapolation of state-change events: Specifically, we predict that, when a linguistic description is shown after a video depicting an object undergoing a state change, the linguistic description in perfective aspect should pull the representation of a video in progress further towards completion, resulting in a bigger representational momentum effect compared to an imperfective description.

Here, we use the perfective/imperfective divide, but unlike previous studies (Madden & Zwaan 2003; Misersky et al., 2021; Minor et al., 2022), we use sentences in the present perfect ("The log has burnt") instead of sentences in the simple past ("The log burnt") for the perfective framing. This choice is motivated by the fact that the English perfect, but not necessarily the English past, highlights the resultant state of the object in state-change events (Kamp & Reyle 1993; Moens 1987; Moens & Steedman 1988; Mittwoch 2008; Parsons 1990). Furthermore, and unlike previous studies, we present the linguistic descriptions of state-change events in intransitive frames to control for potential effects of agentive subjects (Demirdache & Martin, 2015).

We specifically asked whether the use of grammatical aspect modulates the representational momentum effect in state-change events reported in Hafri et al. (2022). An overview of the current studies and their purpose is given in Table 1. First, we replicated Hafri et al's (2022) work in three experiments: Experiment 1a kept the frame rate of videos consistent with Hafri et al. (2022) for a direct replication. Experiment 1b allowed us to explore the sensitivity of representational momentum to the speed of state-change events (Hubbard & Bharucha, 1988; Merz et al., 2019). Experiment 1c was a subset of Experiment 1a that served as a baseline for subsequent experiments.

Then, to understand how people's interpretations of aspect map onto the video stimuli, we conducted two control experiments. Experiment 2a used a two-alternative forced-choice design that asked participants to map sentences to videos; Experiment 2b used only sentences, without showing the video before, to understand how people would construe those events without visual memory.

Lastly, to understand whether the use of grammatical aspect modulates the representational momentum effect in state-change events, we conducted two experiments. In both Experiment 3a and 3b, after watching animation, participants read linguistic descriptions either in perfective or imperfective aspect, except that the timing of the post-event stimulus (linguistic description or visual mask) was controlled in Experiment 3b, consistent with Hafri et al. (2022).

Table 1
A summary table of all experiments. Here and throughout, experiments are color-coded depending on whether they involved language (red) or not (blue).

Experiment	Description	Purpose
Experiment 1a	replication, slow video	to replicate representational momentum effect in state-change events (exact replication of Hafri et al., 2022)
Experiment 1b	replication, fast video	to test video speed effects on state-change representational momentum effect
Experiment 1c	self-replication, slow video, pause frame 60	to rule out regression-to-mean effects from Exp 1a.
Experiment 2a	aspect calibration study	to test how people map our visual stimuli onto their interpretations of grammatical aspect using binary forced choice
Experiment 2b	sentence-frame matching task	to test how grammatical aspect shapes event construal using slider task
Experiment 3a	language-framed, longer delay, pause frame 60	to test how language and the representational momentum effect interact in state-change events
Experiment 3b	language-framed, shorter delay, pause frame 60	to test how language and the representational momentum effect interact in state-change events; to test effect of delay

Experiments 1a-c: Replications of Hafri et al. (2022)

Experiments 1a-c replicated Hafri et al.'s (2022) Experiment 1, which found a representational momentum effect in state-change events. Experiment 1a (Replication, Slow Video) kept the frame rate of videos consistent with Hafri et al. (2022) for a direct replication, while Experiment 1b (Replication, Fast Video) allowed us to explore the sensitivity of representational momentum to the speed of state-change events. Experiment 1c (Self-replication, Slow Video, Pause Frame 60) was a self-replication of Experiment 1a, with a key modification: Participants only viewed videos that paused at frame 60. This design isolated the critical pause frame condition to rule out potential regression-to-the-mean effects introduced by repeated pause frames in Experiment 1a. It also established a clean baseline for subsequent comparisons between Experiments 1c and 3b.

Method

Data availability

All data, code, and stimuli for these experiments are available under https://osf.io/6hkfj.

Participants

For consistency, we replicated the sample size used by Hafri et al. (2022). However, unlike Hafri et al. (2002), who collected 50 but analyzed only 43 participants' data, we decided to analyze 50 full datasets in each study. To reach this sample size, eventually, 80 participants for Experiment 1a, 82 participants for Experiment 1b, and 72 participants for Experiment 1c were recruited from the online platform (Peer et al., 2017; Peirce et al., 2019).

In line with Hafri et al.'s (2022) work, we excluded participants if their response frames, averaged across items, did not monotonically increase with the last frame condition's choice. One participant was excluded in Experiment 1b according to this criterion. Additionally, 30 participants from Experiment 1a, 31 from Experiment 1b and 22 from Experiment 1c were excluded due to a technical issue where the reported frame rate's rounded value (recorded by the browser) deviated from the target of 60 frames per second. All exclusion criteria can be found in our pre-registration: https://osf.io/6hkfj/?view_only = 418a26969bf34f0798f756db32c9d56e.

Stimuli and procedure

We used the same stimuli as Hafri et al. (2022). The state-change videos were animations generated in the 3D modeling software Blender (Version 2.82; https://www.blender.org; Blender Foundation, 2020). Experiments were built using PsychoPy (Peirce, 2007; Peirce et al., 2019).

Each experiment included five state-change videos depicting different objects undergoing physical changes: ice melting, a grape shriveling, a cigarette smoldering, butter deforming and a log burning. Each animation consisted of 240 frames and was paused at one of three frames, 60, 120 and 180 (randomized order, once for each state change) in Experiment 1a and 1b, yielding 15 trials. In Experiment 1c, each animation paused at frame 60, resulting in only 5 trials per participant.

The videos were shown at two different speeds across experiments (Experiment 1a/1c: 30 frames per second, full video duration: 8 s; Experiment 1b: 60 frames per second, full video duration: 4 s). Experiment 1a and 1c maintained the frame rate of videos used by Hafri et al. (2022) to achieve direct replications, whereas the faster speed in Experiment 1b allowed us to explore the sensitivity of representational momentum to the speed of state-change events, as had previously been shown for extended motion events (Hubbard & Bharucha, 1988; Merz et al., 2019).

In each trial (Fig. 2A), participants viewed an animation of one of the state changes, which was stopped before completion and masked for 1,000 ms with a box-scrambled mask. Participants then identified the last frame of the animation that they saw by using a slider that stepped through the animation frame by frame (Fig. 2B). Importantly, the starting position of the slider was fully randomized on each trial. The left end of the slider represented the beginning of the animation, and the right end represented the end of the animation (100% of the 240 frames), which, notably, participants had never seen. When satisfied with their choice, participants clicked a button to move on to the next trial (for a demo video: https://osf.io/6hkfj/?view_only = 418a26969bf34f0798f756db32c9d56e). To become familiar with using the slider, participants first completed a practice trial during the instruction phase. In this trial, they moved the slider back and forth until a button appeared. Participants could not proceed with the experiment until they performed this trial as instructed.

Our dependent variable was the frame error produced by participants' choices of slider position (Fig. 2B): For each trial, we calculated the difference between the chosen frame and the actual pause frame. For instance, if the pause frame was frame 120, a chosen slider position at frame 129 would be a frame error of + 9. We predicted that an overall positive frame error should be observed, indicating participants would remember the last frame they saw as being further forward in time than it actually was.

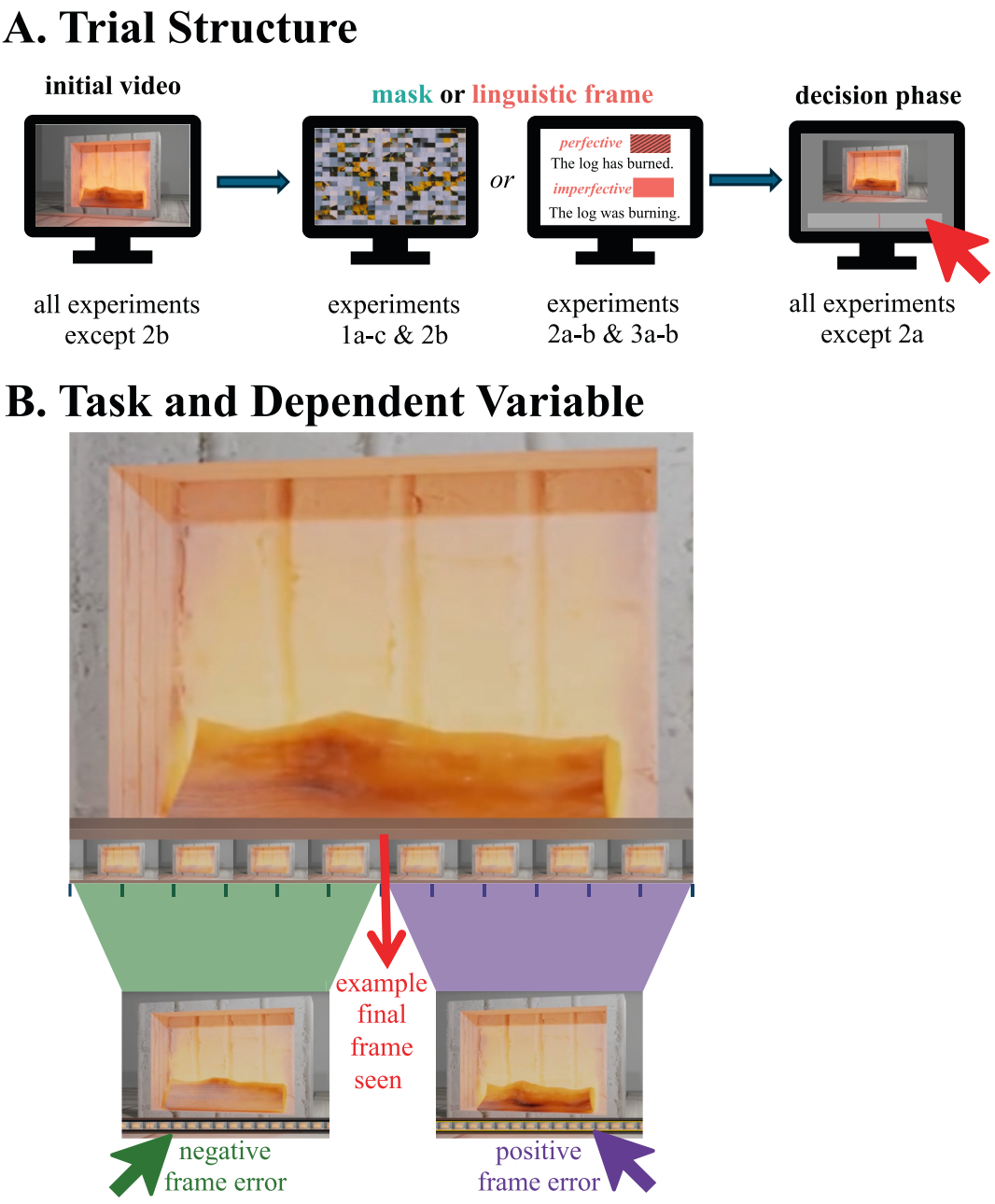

Fig. 2. The basic logic of our experiments. **A**: Basic task design is common to most of our experiments. Participants saw a video, then either a visual mask or a linguistic cue, and then they pulled a slider which stepped through the animation frame by frame to indicate the last frame they had seen from the video. B. Decision phase: Our Dependent Variable was the frame error introduced in the decision phase. If the video stopped, for instance, at the red line, when reconstructing the video, any frame after the last seen indicates a positive frame error (purple), any before, a negative frame error (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

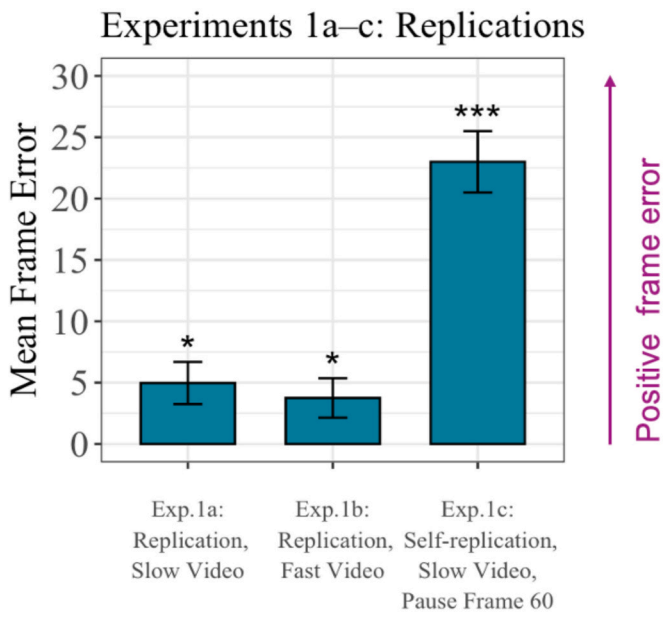

Fig. 3. Results of Experiment 1a (30 frames per second, full video duration: 8 s), Experiment 1b (60 frames per second, full video duration: 6 s), and Experiment 1c (Self-replication, Slow Video, Pause Frame 60 only), showing significant positive mean frame errors. Error bars indicate the standard error of the mean. Asterisks denote statistically significant deviations from zero ($p < .05$) in Experiments 1a and 1b and statistically significant deviations from zero ($p < .001$) in Experiment 1c.

Analyses

In our analyses, we closely followedHafri et al. (2022).[1] To examine the main hypothesis, we conducted a two-tailed, one-sample *t*-test on the frame error averaged across participants against the baseline of the correct pause frame. For secondary analysis, we performed an ANOVA to test how frame error varied by state change, pause frame, and their interaction. We also conducted post-hoc *t*-tests to ask whether the positive frame error remained significant regardless of the pause frame. Preregistration for all analyses is available at: https://osf.io/6hkfj/?view_only = 418a26969bf34f0798f756db32c9d56e.

Results

We observed an overall positive frame error in all three experiments (Fig. 3). In Experiment 1a, the mean frame error was 4.97, indicating that participants reported the remembered frame slightly forward in time relative to the true final frame. This positive shift was significant, replicating the representational momentum effect reported by Hafri et al. (2022) ($t(49) = 2.41$, $p < .02$, $d = 0.34$, 95% CI [0.82, 9.11]). In Experiment 1b, we again found a significant positive frame error ($M = 3.75$; $t(49) = 2.32$, $p < .03$, $d = 0.33$, 95% CI [0.50, 7.00]), confirming the robustness of this effect across replications. Experiment 1c, which included only the pause-at-frame-60 baseline condition for later comparison with Experiment 3b, also showed a robust positive representational momentum effect ($M = 23$; $t(49) = 7.64$, $p < .001$, $d = 1.08$, 95% CI [17.24, 29.56])).

Secondary analyses showed that the effect was not consistent across all items or pause frame conditions: in 1a, three of five scenarios and one pause condition showed significant forward errors, while in 1b, two of five scenarios and two of three pause conditions did (see Appendix Tables A1–A2).

Further, there was no significant difference of mean frame error between Experiment 1a and 1b ($t(1487.1) = 0.73$, $p = .46$, 95% CI [–2.04, 4.48]), suggesting that the speed of the state change did not modulate the magnitude of representational momentum.

Discussion of experiments 1a-c

To summarize, the results of Experiments 1a-c revealed a significant representational momentum effect for state-change events, establishing the baseline for our subsequent linguistic manipulations. Specifically, the positive frame errors in all three replications indicate that people indeed mentally extrapolated state-change events and objects' states were misremembered being more changed forward in time.

Experiments 1a-b used the same design and protocol as in Hafri et al. (2022), but the magnitude of the representational momentum effect was smaller than the positive mean frame error of 13.25 reported by Hafri et al. (2022). We believe this difference to be the result of natural variation, which has been observed elsewhere (e.g., Binder, 2022; Casey et al., 2017; Goodman et al., 2013), since we used Hafri et al.'s (2022) code, stimuli, and even mode of data collection.

Furthermore, we found no significant difference in mean frame error between Experiments 1a and 1b. This suggests that the speed of objects' state changing did not affect the magnitude of mental extrapolation. We will return to this point in the General Discussion.

By including only the pause-at-frame-60 condition, Experiment 1c aimed to address a potential methodological artifact in Experiment 1a, where multiple pause-frame conditions might have introduced regression-to-the-mean effects.[2] The results replicated the key finding from Experiment 1a: Participants remembered objects in state-change events as being more changed than they actually were. This provides evidence that the representational momentum effect for dynamic state changes is robust and not dependent on the inclusion of multiple pause frame conditions.

Experiments 2a-b: Understanding the effect of aspect on video stimuli

To understand how people's interpretations of aspect map onto the video stimuli, we conducted two control experiments. Experiment 2a employed a two-alternative forced-choice design in which participants matched sentences to videos. Experiment 2b presented only sentences, without before video exposure, to assess how participants construed the described events in the absence of visual memory.

Experiment 2a: Aspect calibration study

Most previous work (e.g., Kang et al., 2020; Madden & Zwaan, 2003) used sentence-picture verification tasks to explore how people mentally represent and integrate aspectual information with static images, but not animations. Experiment 2a used a two-alternative forced-choice task to understand how the videos used correspond to people's interpretation of grammatical aspect.

Method

Data availability

All data, code, and stimuli for these experiments are available under https://osf.io/t5whc/?view_only = 7b868451ae7144d383ed3843d023b410.

Participants

To determine the required sample size, we conducted a power analysis using the *simr* package (Green & Macleod, 2016) in R based on

[1] While we used simple *t*-tests and ANOVAs in Experiment 1a, 1b and 1c to keep in line with Hafri et al.'s 2022 analysis, later experiments were analyzed by mixed-effects models.

[2] In addition, we conducted supplementary analyses that revealed no regression-to-the-mean effects in Experiment 1a (see "Supplementary Analyses", https://osf.io/6hkfj/files/4etjp).

pilot data. Results indicated that 30 participants would result in 100% power for detecting an effect, at a significance criterion of α = .05. To analyze 30 full datasets, 50 English native adults were recruited via Prolific (Peer et al., 2017; Peirce et al., 2019). We excluded any participant who did not contribute to a complete data set. We also excluded participants whose reported frame rate's rounded value (recorded by the browser) deviated from the target of 60 frames per second. 20 participants were excluded due to this technical issue. The exclusion criteria can be found in our pre-registration: https://osf.io/t5whc/?view_only = 7b868451ae7144d383ed3843d023b410.

Stimuli and the procedure

In this study, participants watched each of Hafri et al.'s (2022) videos, stopping at either pause frame 60, 120, 180 or 240 (randomized order, once for each state change), resulting in 20 trials. These stimuli covered all stages of an unfolding state-change event, including the end of the videos, corresponding to different states of the affected object. This extends most previous sentence-picture verification work (Kang et al., 2020; Madden & Zwaan, 2003; Misersky et al., 2021), which typically presented only static images depicting either a fully completed or an ongoing action, rather than capturing the dynamic progression of change.

In a two-alternative forced choice task, participants were asked to select the sentence that better described the event: Either a sentence in the perfective aspect (e.g. *The* log *has burnt*), or a sentence in the imperfective aspect (e.g. *The* log *was burning*). For each trial, we recorded participants' choices, and our dependent variable was the choice of the perfective sentence.

Analyses

We built a linear mixed-effect regression model with choice predicted by pause frame and with random intercepts for item and participant. Significance was assessed by a likelihood ratio test that compared the full model with a null model, which excluded our predictor. As for exploratory analysis, we performed pairwise comparisons of the pause frames. Preregistration for all analyses is available at: https://osf.io/6hkfj.

Results

As shown in Fig. 4, the imperfective sentence was always a better description, but the perfective sentence became more acceptable as the video progressed—especially after the halfway point—corresponding to the affected object approaching its endpoint of state change. This pattern was confirmed statistically: A model comparison showed that pause frame had a significant effect on choice ($\chi 2(3) = 59.79$, $p < 0.001$).

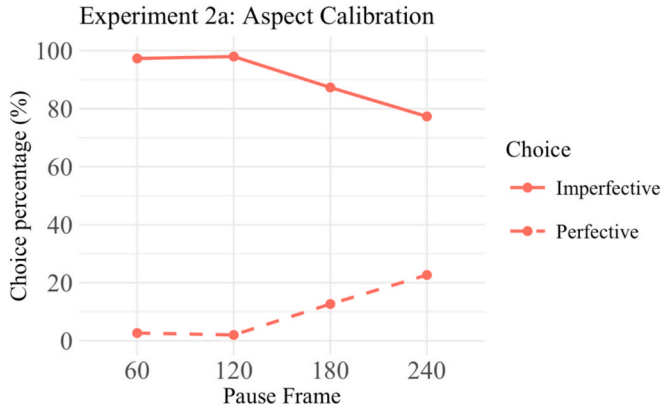

Fig. 4. Results of Experiment 2a (Aspect Calibration Study), showing participants' choice percentage of perfective and imperfective sentences across four pause frames.

To further understand this association across four pause frames, we ran two linear mixed-effect regression models. The first model took pause frame 60 as the reference level, and the second model took pause frame 180 as the reference level. In both models, the predictor was the pause frame, and the outcome was the choice, in which perfective sentence was coded 1. With the intercept at frame 60, the model showed no significant difference to frame 120 (β = -0.32, z = -0.40, p = .69, 95% CI [-1.89, 1.25]). However, there was a significant difference to frame 180 (β = 1.99, z = 3.27, p = .001, 95% CI [0.80, 3.18]) and 240 (β = 2.97, z = 4.93, p < .001, 95% CI [1.79, 4.16]). When the reference level was frame 180, the model showed significant differences to all other pause frames and crucially, the effect of pause frame 240 was significant (β = 0.99, z = 4.93, p = 0.009, 95% CI [0.25, 1.72]). This further indicates that when the video unfolded, especially after the half-way point, people were significantly more likely to map the video up to the last frame seen to a perfective sentence.

Discussion of Experiment 2a

Taken together, these results indicate that people's interpretations of grammatical aspect map onto the video stimuli used: The choice of an imperfective sentence (e.g., "The log was burning") dominated throughout the video, and importantly, the choice of a perfective sentence (e.g., "The log has burnt") increased as the video unfolded.

This aligns with previous work showing grammatical aspect provides an internal temporal perspective on an event, with imperfective aspect mapping onto events in progress without specifying their limits, whereas perfective aspect mapping onto events with completed boundaries (Madden & Zwaan, 2003; Misersky et al., 2021; Magliano & Schleich, 2000). Based on this theoretical foundation and the result of Experiment 2a, we predicted that when participants read a perfective sentence after a progressive video, the perfective sentence should pull the representations of the in-progress video further along. This is what we tested in Experiment 3a-b.

Experiment 2b: Sentence-frame matching task

While Experiment 2a confirmed the standardly assumed mapping of imperfective and perfective aspect through different stages of an event, Experiment 2b was designed to understand how participants would construe the last frame of a video based on linguistic cues alone: If we find that linguistic framing modulates the representational momentum effect, one way to explain this pattern would be to assume that participants simply ignore the video, and reconstruct the last frame based on linguistic input alone. This explanation would entail that there is no interaction between linguistic framing and a visual memory trace. To prevent this interpretation of any subsequent results, we conducted a version of Hafri et al. (2022), but instead of the video, only a linguistic cue preceded the decision phase.

In that regard, Experiment 2b served as a norming/matching benchmark for how language alone guides frame selection.

Method

Participants

To reach a full dataset of 200 participants, we recruited 320 English native speakers from Prolific (Peer et al., 2017; Peirce et al., 2019); and 120 were excluded because their frame rates' rounded value (recorded by the browser) deviated from the target of 60 frames per second. We recruited more participants than in the previously reported studies to increase power, given the subtlety of the slider task combined with the linguistic manipulation introduced here.

Stimuli and procedure

The stimuli and design were identical to those used in Hafri et al. (2022), except that here, participants did not see the video of an event. Instead, they saw a visual mask (2 s) before reading a sentence (1 s) and

completing the slider task. Participants were instructed to pull the slider to create a scene that best matched the meaning of the sentence they had just read.

Analyses and results

In this study, the Dependent Variable was the Response Frame, that is, the frame selected in the absence of any event encoding. Across conditions, the mean response frame was 180.65 (SD = 68.43), indicating that participants' memory judgments corresponded to a relatively late stage of the event (Fig. 5).

A linear mixed-effect model (estimated using ML and the BOBYQA optimizer) with random intercepts for participant and item revealed a robust effect of Aspect: Responses following perfective descriptions (M = 229.4, SD = 28.33) were significantly later than those following imperfective descriptions (M = 131.6, SD = 61.36; $t(1988) = 50.46$, $p < .001$, 95% CI [94.06, 101.66]).

Discussion of Experiment 2b

Experiment 2b was designed as a control to determine whether any effects observed in the subsequent linguistic experiments (Experiments 3a-b) could be attributed solely to linguistic framing. In this study, participants construed the last frame of an event without ever seeing the video, relying exclusively on linguistic cues.

The results showed a robust effect of grammatical aspect in the absence of any visual information: Participants selected significantly later frames after perfective sentences than after imperfective sentences. This finding demonstrates that linguistic framing alone can bias participants' mental representations of an event's temporal stage.

Experiment 2b serves as a norming study and its outcome provides a crucial interpretive baseline for understanding the results of the following linguistic experiments. If there is no interaction between linguistic framing and a visual memory trace, participants' responses in the subsequent linguistic experiments should resemble the pattern observed in Experiment 2b. Their final judgments should be determined primarily by linguistic cues, unaffected by how the event was visually encoded. However, any deviation from this linguistic-only pattern would indicate that participants are not ignoring the visual input, and that there is an interaction between language and visual encoding of these events.

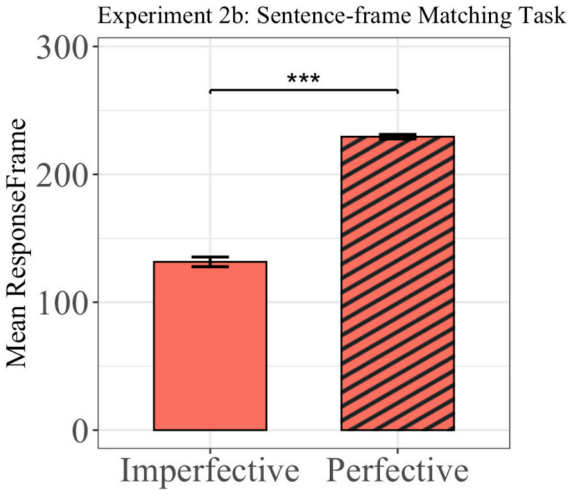

Fig. 5. Results of Experiment 2b (Sentence-frame Matching Task), showing a significant difference between the mean response error induced when people read either imperfective or perfective sentences. Error bars indicate the standard error of the mean. The asterisk denotes a statistically significant difference between the imperfective and perfective conditions ($p < .001$).

Experiment 3a and 3b: Understanding the effect of aspect on state-change representational momentum effect

Experiments 3a and 3b ask our main question: Whether linguistic framing influences the fundamental cognitive processes of mental extrapolation in state-change events, specifically, whether the use of grammatical aspect modulates the representational momentum effect found by Hafri et al. (2022). Here, we focused on the perfective/imperfective divide, using the present perfect ("The log has burnt") in the perfective condition. Additionally, we used intransitive frames to control agent effects on culminating construal. Based on results from Experiment 2, we predicted that when language is introduced after a video depicting an object undergoing a state change, the sentence in the perfective aspect should pull the representation of the progressive video further, resulting in a larger effect of representational momentum than in the imperfective condition.

Experiments 3a and 3b used the same design and procedure, differing only in the timing of the linguistic description. In Experiment 3a, the sentence was presented for 3,000 ms before the event unfolded in the video, whereas in Experiment 3b, it was shown for 1,000 ms, matching the timing of the visual mask used by Hafri et al. (2022). By aligning the post-event stimulus timing in Experiment 3b with Hafri et al.'s study, we aimed to rule out the possibility that any effects observed in Experiment 3a were due to the longer processing time afforded by the linguistic description. Furthermore, comparing Experiment 3b with Experiment 1c (Self-replication, Slow Video, Pause Frame 60) allows us to assess whether linguistic encoding produces a general influence on overall memory accuracy.

Method

Data availability

All data, code, and stimuli for these experiments are available under https://osf.io/w4yeq/ (Exp3a) and https://osf.io/w5a42 (Exp 3b).

Participants

To determine the required sample size, we conducted a power analysis using the *simr* package (Green & Macleod, 2016) in R based on pilot data. Results indicated that the required sample size to achieve 80% power for detecting an effect, at a significance criterion of $\alpha = .05$, was 200. To analyze 200 full datasets, 333 English native adults were recruited in Experiment 3a and 306 were recruited in Experiment 3b via Prolific (Peer et al., 2017; Peirce et al., 2019). We excluded any participant who did not contribute a complete data set. Again, we excluded any participant whose reaction times are shorter than 2 s and longer than 15 s in more than 1/3 of trials, and we also excluded participants whose reported frame rate's rounded value (recorded by the browser) deviated from the target of 60 frames per second. 133 participants in Experiment 3a and 106 participants in Experiment 3b were excluded due to this technical issue. The exclusion criteria can be found in our pre-registrations.

Stimuli and procedure

For the first replication, we adapted the protocol from Hafri et al. (2022). In Experiments 3a and 3b, the only change compared to Experiment 1a was that after each video, people saw a sentence describing the event using either perfective or imperfective aspect (See Fig. 2A). For example, after watching a video depicting a log burning, participants read a sentence in either perfective aspect ("*The log has burnt.*") or imperfective aspect ("*The log was burning.*"). The sentence was presented for 3 seconds in Experiment 3a and 1,000 ms in Experiment 3b.

In both studies, we used a pause frame of 60, with video durations fixed at 2 s. This frame was chosen because it was least likely to affect the telicity of the event and did not constrain the slider responses near the end. Furthermore, in both Hafri et al.'s (2022) study and our

replications, pause frame 60 produced the most stable effects, making it ideal for testing the impact of aspect. All participants viewed each state-change event (melting, deforming, burning, shriveling, and smoldering) in both perfective and imperfective conditions, resulting in 30 trials.

Analyses

In both Experiments 3a and 3b, we built a mixed-effect regression model with frame error predicted by aspect, and with random intercepts for item and participant. Exploratory analyses followed the approach of Hafri et al. (2022), who performed two-tailed *t*-tests for state change; here, we applied the same logic to grammatical aspect, comparing each condition against the last frame seen (baseline).

As for the between-study comparison, we built another linear mixed-effect regression model with Frame Error (DV) predicted by delay time (1-second or 3-second), and with random intercepts for item and participant. Exploratory analyses followed Experiment 3a and 3b.

Results

Across both experiments, we replicated the representational momentum effect, reflected by a significant overall positive frame error (Experiment 3a, M = 7.4, $t(1612) = 18.24$, $p < .001$; Experiment 3b, M = 5.35, $t(1608) = 2.25$, $p < .03$). The results indicated that participants reported a frame error forward in time relative to the true final frame in both experiments.

Further, as predicted, in Experiment 3a, sentences in the perfective aspect led people to reconstruct an event that was further advanced (M = 8.3) than sentences in the imperfective aspect (M = 6.5), a statistically significant difference ($t(1608) = 2.25$, $p < .03$, 95% CI [0.21, 3.11]). A similar pattern emerged in Experiment 3b, where sentences in the perfective aspect again led to more advanced event reconstructions (M = 6.1) than sentences in the imperfective aspect (M = 4.4), a difference that was statistically significant ($t(1638) = 2.11$, $p = .035$, 95% CI [0.11, 2.88]). Together, these results confirm that grammatical aspect reliably modulates mental extrapolation of state-change events, independent of the timing of linguistic presentation.

The between-study comparison showed a significant main effect of delay time ($t(3249) = 2.06$, $p = 0.04$, 95% CI [5.54e-03, 0.23]), and participants in Experiment 3a with linguistic descriptions showing for 3,000 ms reported a larger frame error. However, the interaction between delay time and grammatical aspect was not significant ($t(3249) = 0.21$, $p = 0.831$, 95% CI [-0.11, 0.14]), indicating that the effect of aspect did not depend on delay time.

Discussion of Experiment 3a and 3b

Experiments 3a and 3b examined whether linguistic framing influences the process of recalling mentally extrapolated state-change events and whether this effect depends on the timing of linguistic input.

In Experiment 3a, participants who read sentences in the perfective aspect reconstructed the event as more advanced than those who read sentences in the imperfective aspect, demonstrating that grammatical aspect modulates event memory. Experiment 3b replicated this aspect effect while controlling the timing of sentence presentation to match the mask duration used by Hafri et al. (2022). The persistence of the effect under this stricter timing condition confirms that linguistic framing influences mental extrapolation independently of processing time, strengthening the interpretation that high-level semantic encoding can modulate the perceptual reconstruction of dynamic events.

These findings extend existing evidence on how linguistic framing, particularly grammatical aspect, affects visual event representations and align with broader research on the interaction between language and perception (e.g., Forder & Lupyan, 2019; Lupyan et al., 2020; Maier & Abdel Rahman, 2019; Reed & Vinson, 1996).

Between-Experiment Comparison: Self-replication, Slow Video, pause frame 60 (Exp 1c) / Language-framed, shorter delay, pause frame 60 (Exp 3b)

To assess whether linguistic encoding exerts a general effect on overall memory accuracy, we conducted a between-experiment comparison. Here, we compared data from Experiment 1c, which included only the pause-at-frame-60 condition, with data from Experiment 3b (Fig. 6), which also had only the pause-at-frame-60 condition but presented a linguistic description at the same 1000 ms delay as in Experiment 1a (thus replicating Hafri et al.'s original timing parameters).

We conducted a linear mixed-effects regression predicting frame error from experiment condition (Experiment 1c: non-linguistic versus Experiment 3b: linguistic), with random intercepts for participant and item. The analysis revealed a robust and statistically significant effect of language ($t(1888) = -9.80$, $p < .001$, 95% CI [-21.07, −14.05]). As shown in Fig. 7, frame error was substantially smaller in Experiment 3b (M = 5.84, 95% CI [0.30, 11.40]) compared to Experiment 1c (M = 23.40, 95% CI [17.69, 29.10]). This indicates that participants were considerably more accurate in recalling the final frame of the state-change videos when a linguistic description accompanied the visual encoding, relative to when no language was provided.

General Discussion

This paper set out to understand whether linguistic framing, especially the use of grammatical aspect, could modulate the effect of representational momentum in state-change events. We predicted that descriptions of videos of state-change in perfective aspect would result in a bigger representational momentum effect, compared to descriptions in imperfective aspect.

We first replicated Hafri et al.'s (2022) work in three experiments: Experiment 1a kept the frame rate of videos consistent with Hafri et al. (2022) for a direct replication. Experiment 1b examined how the representational momentum effect responds to the speed of state-change events. Experiment 1c, a subset of Experiment 1a, served as a baseline for subsequent experiments. In all three replications, we found a robust representational momentum effect.

Next, to understand how people's interpretations of aspect map onto our video stimuli, we conducted two control experiments. Experiment 2a used a two-alternative forced choice design where participants matched sentences to videos; Experiment 2b presented only sentences, without videos, to understand how people would construe those events without visual memory. In both experiments, we found a reliable mapping between less advanced events to imperfective aspect and more advanced effects to perfective aspect.

Finally, to understand whether the use of grammatical aspect modulates the representational momentum effect in state-change events, we conducted two experiments. In both Experiment 3a and 3b, after

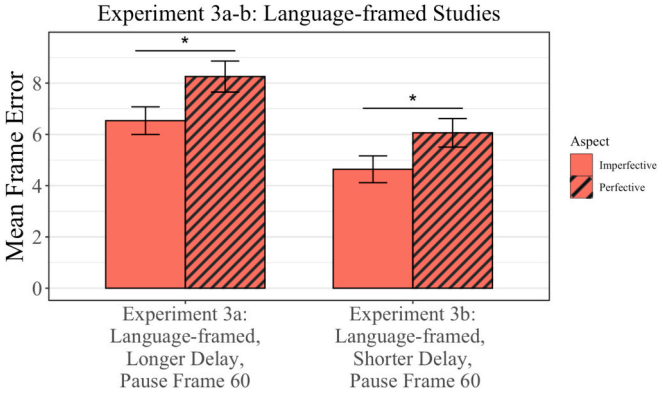

Fig. 6. Results of Experiment 3a (Language-framed, Longer Delay, Pause Frame 60) and Experiment 3b (Language-framed, Shorter Delay, Pause Frame 60), showing a significant difference between the mean frame error induced when people read either imperfective or perfective sentences. Error bars indicate the standard error of the mean. The asterisk denotes a statistically significant difference between the imperfective and perfective conditions ($p < .05$).

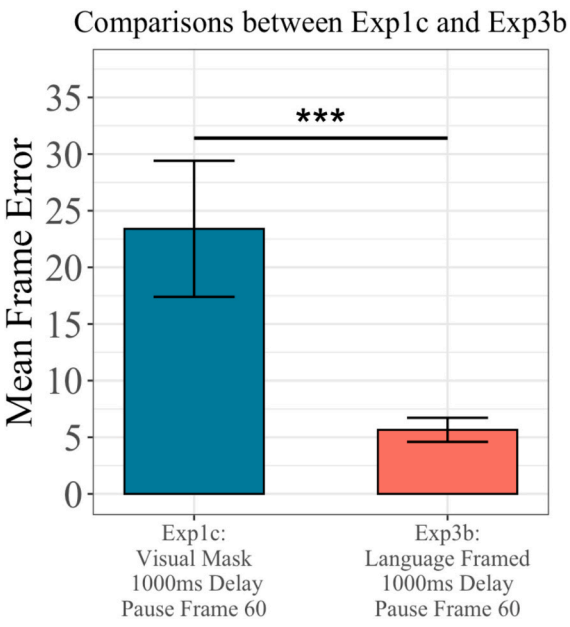

Fig. 7. Results of between-experiment comparison, showing a significant difference in mean frame error between Experiment 1c (Visual Mask, 1000 ms Delay, Pause Frame 60) and Experiment 3b (Language Framed, 1000 ms Delay, Pause Frame 60). Error bars indicate the standard error of the mean. The asterisk denotes a statistically significant difference between conditions ($p < .001$).

watching animations, participants read linguistic descriptions either in perfective or imperfective aspect, except that the timing of the post-event stimulus (linguistic description or visual mask) was controlled in Experiment 3b, consistent with Hafri et al. (2022). As predicted, perfective aspect led to a larger representational momentum effect than imperfective aspect.

The effect of grammatical aspect on the representational momentum effect

We found that grammatical aspect affects the representational momentum effect, which was greater after sentences in perfective than imperfective aspect in both Experiment 3a and 3b. Taken together, these findings suggest that the representational momentum effect observed in state-change events is not the result of a fully encapsulated process. Instead, it is at one point sensitive to linguistic framing, specifically the use of grammatical aspect. This linguistic influence reveals that representations of ongoing change events are not confined to bottom-up visual processing but are accessible to and shaped by higher-level conceptual input. Whether the effect is post-hoc or pre-processing is outside the scope of this paper.

This closely aligns with two claims in the Intersecting Object History theory of event representation (Altmann & Ekves, 2019; also see Hindy et al., 2012; Kang et al., 2020; Prystauka et al., 2024; Solomon et al., 2015). This theory first and foremost proposes that event understanding, both linguistic and non-linguistic, is built upon dynamic representations of intersecting object histories. However, evidence for this claim has come mostly from static object representations, such as pictures of its beginning and end states, not dynamic videos. What our study, but also the study of Hafri et al. (2022), provides, is not only evidence for the representational momentum effect, but also evidence that people track objects' histories: In order for the representational momentum effect to occur, the objects' physical state must be tracked as they undergo a change, just as the Intersecting Object History account suggests.

A second claim of the Intersecting Object History theory is that during language comprehension, objects' histories – their evolving state changes – are actively construed (Altmann & Ekves, 2019; Hindy et al., 2012; Kang et al., 2020). During construal, aspectual manipulations provide a guideline to the comprehender as to how far the change of an object must be tracked and construed, and how this interacts with the actual visual input through representational interfaces (e.g., Ferreira & Barker, 2024; Jackendoff, 2023). Our Experiments 3a and 3b provide empirical support for this prediction, demonstrating that aspectual manipulations systematically influence how comprehenders interpret and remember state changes. In addition, our results are in line with previous work (Feist & Gentner, 2007; Gentner & Loftus, 1979), suggesting that verbalizing an event while also visually processing it can influence memory for that event. To our knowledge, our study is the first work using continuous dynamic stimuli showing state-change events, showing the effect of grammatical aspect on event memory.

A long tradition in the literature reports a goal/endpoint advantage: endpoints are frequently prioritized over starting points in language and, under some conditions, in memory (e.g., Lakusta & Landau, 2005, 2012; Lakusta & Carey, 2015; Papafragou, 2010; Regier & Zheng, 2007), although this bias seems to be somewhat context-dependent, not universal. For instance, Lakusta and Landau (2012) showed that adults and children display a goal bias for both agentive and non-agentive events in linguistic tasks, but in non-linguistic memory the bias emerges primarily for agentive, goal-directed motion. However, taken together, previous findings motivate a prediction for our paradigm: If language simply heightens endpoint focus, linguistic cues should increase forward displacement (RM).

Against this backdrop, it is useful to distinguish telicity, a semantic property of predicates that encode whether an event has a natural endpoint (e.g., "the butter melts (in ten minutes)" vs. "the butter sits on the plate (#in ten minutes)"). Telic predicates refer to events of change with inherent endpoints, whether it is the building of something that didn't exist before (coming into existence), or the melting of something that existed in a different form (change of state): Hence their compatibility with modifiers of the form in X time, that quantify how long it takes to achieve the end state. By contrast, atelic predicates refer to states or activities without an inherent endpoint or end state, such as walk around the city or siting by the window, and these predicates are generally incompatible with those same modifiers. One might expect that telic verbs, by highlighting event boundaries, could align with or even amplify the cognitive bias for goal states and endpoints.

However, our findings complicate this assumption. Specifically, we observed that the effect of representational momentum was reduced in the presence of telic linguistic descriptions, contrary to what would be expected if telicity merely intensified goal bias. If telicity alone heightened attention to endpoints, participants might over-anticipate change, leading to a larger effect of representational momentum, not smaller, closer to the pattern observed in Experiment 2b. Our results instead suggest that grammatical aspect, not telicity alone, influences how object change is mentally tracked beyond the influence of telicity alone.

What is important to reiterate is that we also found, across different designs, that people's interpretations of aspect mapped onto our video stimuli. Experiment 2a employed a two-alternative forced-choice design in which participants matched sentences to videos. Unlike previous work on grammatical aspect using static images, which depicted an object either undergoing no state-change or substantial state-change (Kang et al., 2020; Madden & Zwaan, 2003; Misersky et al., 2021), our work captures the dynamic progression of change by including different stages of an unfolding state change. We found that the sentence in imperfective aspect was generally preferred throughout the video, but the sentence in perfective aspect became more acceptable as the video approached its endpoint. Not surprisingly, in Experiment 2a, participants did not predominantly select the sentence in perfective aspect even after viewing the full event, possibly because the event endpoints were not fully realized in these animations (e.g., the log did not completely transform into ashes), but perfective aspect was selected more in more progressed change-of-state events.

While Experiment 2a confirmed the standardly assumed mapping of

imperfective and perfective aspect through different stages of an event, Experiment 2b conducted a version of Hafri et al. (2022): Without exposure to videos, participants read sentences and used a slider displaying a continuous animation to indicate how they mentally represented the described events. This design served as a norming task and allowed us to capture how grammatical aspect alone, without direct visual input, shapes people's conceptualization of ongoing versus completed changes.

Connected to this, one way to interpret our results would be to argue that participants simply ignored the video and reconstructed the last frame based on linguistic input alone. However, Experiment 2b provides a baseline demonstrating that such an explanation cannot account for the current findings. In Experiment 2b, participants made judgments based only on linguistic information, without visual input. If linguistic framing alone had driven participants' responses in Experiments 3a and 3b, their pattern of results should have resembled that of Experiment 2b. Instead, we found reduced representational momentum effects in both Experiments 3a and 3b, indicating that participants did not disregard the visual input. These results suggest that the observed effects arose from an interaction between linguistic framing and visual encoding.[3]

In sum, consistent with prior psycholinguistics work, both experiments showed that grammatical aspect maps onto events in a systematic and predictable way, specifically, perfective aspect maps onto completed events, reinforcing a focus on the final state of affected objects (Madden & Zwaan, 2003; Misersky et al., 2021). In addition to confirming effects of perfective aspect, our findings are in line with previous work showing that imperfective aspect functions as a "lens" into the internal dynamics of an event —- specifically by directing attention to ongoing changes in an object's state. This focus on dynamic object states goes beyond other types of event-internal information, such as participants, instruments, locations, or sub-events (Bevacqua et al., 2021; Carreiras et al., 1997; Ferretti et al., 2007; Morrow, 1985; Morrow, 1990; Madden & Therriault, 2009; Truitt & Zwaan, 1997; Wampler & Wittenberg, 2023).

The effect of temporal parameters on the representational momentum effect

In addition to our main question about the influence of linguistic framing on the state-change Representational Momentum effect, this set of studies is also informative for understanding the role of timing in the task. For instance, one open question our baseline studies addressed was the role of speed of change for the Representational Momentum effect. We conducted two versions of Hafri et al.'s (2022) study: Experiment 1a (30 frames per second, full video duration: 8 s) and Experiment 1b (60 frames per second, full video duration: 6 s). In both, replicating Hafri et al. (2022), we found a significant positive representational momentum effect for state-change events. However, in contrast to prior research (e.g., Hubbard & Bharucha, 1988; Merz et al., 2019), the speed of transformation in the video stimuli did not influence the magnitude of this effect.[4] Specifically, participants did not exhibit greater forward displacement in memory in the fast video condition (Experiment 1b) compared to the slow video condition (Experiment 1a). This outcome contrasts with previous studies on representational momentum, demonstrating that faster-moving targets result in greater forward displacement in memory, which has been explained by the 'speed prior' account (Merz et al., 2022). This account proposes that observers rely on internalized expectations about typical object velocities to predict future positions: expected speed shapes cognitive representations of movement, leading to stronger anticipatory displacement for faster motion.

The lack of a speed effect in our studies may stem from differences in event type. While previous studies focused on changes in an object's location or orientation, our study examined state changes (e.g., the state of a burning log), which involve transformations in both internal and external properties. Unlike location changes, state changes are not associated with unidimensional visual cues to velocity, which could reduce participants' sensitivity to speed as a predictive cue. This interpretation aligns with findings from Hafri et al. (2022), who demonstrated that the representational momentum effect can emerge even when participants view a single static frame and indicate change using a slider task. Such results suggest that extrapolation in state-change events does not depend on continuous motion or variation in speed. Similarly, our Experiment 2b (Sentence-frame Matching Task) shows that aspect-driven extrapolation can emerge even without any visual input, reinforcing that speed information is not necessary for state-change representational momentum effect. Future research could further test this possibility by examining whether slowing a state-change event below a certain speed threshold influences the representational momentum effect. One prediction is that participants rely primarily on the final observed state of the object, extrapolating future changes to the same extent regardless of the transformation rate. This would imply that the speed of change does not affect the degree of extrapolation.

A second temporal parameter that may have influenced the Representational Momentum effect was the delay between the end of the video and the response. Comparing Experiment 3a to Experiment 3b, the difference in the delay between the end of the animation and the test produced an independent main effect on the magnitude of the representational momentum effect. Notably, participants who were exposed to a shorter delay time (Experiment 3b) exhibited a smaller representational momentum effect. This pattern suggests that when the linguistic description follows the visual input more immediately, participants can integrate linguistic and visual information more efficiently before the visual memory trace decays. Although linguistic processing occurs rapidly, within a few hundred milliseconds (Bemis & Pylkkänen, 2011; Fallon & Pylkkänen, 2024; Pylkkänen & Marantz, 2003), a longer delay may reduce the overlap between linguistic integration and the visual representation. Our work also suggests that information encountered most recently remains more accessible in working memory (Baddeley & Hitch, 1974; Broadbent & Broadbent, 1981; Davelaar et al., 2005; Rose et al., 2001).

Crucially, for the purposes of our investigation, the effect of grammatical aspect on representational momentum in state-change events persists in both Experiments 3a and 3b, regardless of variation in the presentation time of linguistic description. An additional key finding was that participants were more accurate in recalling the final frame (i.e., showed smaller frame error) when a linguistic description was presented after the video, as compared to trials where only a visual mask was shown (e.g., Experiment 3b vs. Experiment 1a). This result indicates that any linguistic encoding improved overall memory accuracy. This language advantage persisted even when the timing of the post-event stimulus (linguistic description or visual mask) was controlled, consistent with Hafri et al.'s (2002) findings.

We speculate that this overall language advantage stems from what was presented during delay. In our baseline studies (Experiment 1a-b), the visual mask may interfere with visual working memory consolidation of the video (Blalock, 2013; Vogel et al., 2006), which can overwrite or degrade stored information, leading to a greater effect of representational momentum in our case. In contrast, the linguistic description, presented as a sentence in the center of the screen, functioned as a conceptual cue that overlapped semantically with the encoded visual

[3] We acknowledge Reviewer 1's concern that the results of Experiment 2b do not permit a decisive inference that grammatical aspect altered the contents of visual memory per se. We think Experiment 2b serves as an independent motivation to determine whether language alone can account for the effects on RM, or whether it is an interaction of the linguistic framing and the visual encoding.

[4] While it is in principle possible to separate speed of transformation and video duration, they are conflated in telic events. We leave the question whether either transformation speed or duration could be individually impact the representational momentum effect to future research.

event (Craik & Lockhart, 1972; Tulving & Thomson, 1973). By prompting participants to engage with event-relevant information (e.g., whether a log "was burning" or "has burnt"), the linguistic input may reinforce participants' mental representation of the event, making it easier to recall accurately.[5]

Conclusion

Taken together, our work presents evidence that the representational momentum effect in state change events can be influenced by linguistic framing, specifically the use of grammatical aspect. It contributes to a line of research arguing that visual processes are, at one point, also susceptible to top-down conceptual information (e.g., Forder & Lupyan, 2019; Lupyan et al., 2020; Maier & Abdel Rahman, 2019; Reed & Vinson, 1996) through an 'informational conduit between language and the visual system' (Jackendoff, 2023) which links the representations of two distinct representational formats during processing or in memory. Whether our evidence of linguistic framing on the representational momentum effect results from information exchange between language in vision during memory recall, or already during visual event encoding, is a question that should be addressed in future studies.

CRediT authorship contribution statement

Xueyi Yao: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Natalia Jardón:** Writing – review & editing, Supervision. **Jonathan Kominsky:** Writing – review & editing, Supervision, Software. **Eva Wittenberg:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare no financial interests/personal relationships which may be considered as potential competing interests.

Appendix

Table A1
One sample *t*-test statistics across state changes in both Experiment 1a and 1b.

State Change	Mean Frame Error	Statistics	Experiment
Ice melting	M = 8.95	$t(49) = 4.1$, $\mathbf{p < .001}$, $d = 0.57$, 95% CI = [4.51, 13.40]	1a
Log burning	M = 11.42	$t(49) = 2.31$ $\mathbf{p = .03}$, $d = 0.33$, 95% CI = [1.48, 21.36]	1a
Butter deforming	M = −3.68	*t(49) = -1.17, p = 0.37, d = -0.17, 95% CI = [-10.02, 2.66]*	1a
Cigarette smoldering	M = 3.51	*t(49) = 2.32,* $\mathbf{p = .02}$*, d = 0.33, 95% CI = [0.47, 6.55]*	1a
Grape shriveling	M = 4.64	*t(49) = 1,86, p = 0.06, d = 0.26, 95% CI = [-0.37, 9.65]*	1a
Ice melting	M = 7.95	*t(49) = 3.61,* $\mathbf{p < .001}$*, d = 0.51, 95% CI = [3.53, 12.38]*	1b
Log burning	M = 9.86	*t(49) = 2.77,* $\mathbf{p < .001}$*, d = 0.39, 95% CI = [2.71, 17.01]*	1b
Butter deforming	M = -2.75	*t(49) = -0.9, p = 0.37, d = -0.13, 95% CI = [-8.90, 3.39]*	1b
Cigarette smoldering	M = -0.14	*t(49) = -0.1, p = 0.92, d = -0.01, 95% CI = [-3.50, 2.77]*	1b
Grape shriveling	M = 3.82	*t(49) = 1.40, p = 0.17, d = 0.2, 95% CI = [-1.67, 9.32]*	1b

Table A2
One sample *t*-test statistics across pause frames in both Experiment 1a and 1b.

Pause Frame	Mean Frame Error	Statistics	Experiment
60	M = 11.42	$t(49) = 3.78$, $\mathbf{p < .001}$, $d = 0.53$, 95% CI = [5.34, 17.50]	1a
120	M = 3.81	$t(49) = 1.52$, $p < .0.14$, $d = 0.21$, 95% CI = [5.34, 17.50]	1a
180	M = -0.32	*t(49) = -0.13, p = 0.90,* $d = -0.02$*, 95% CI = [5.34, 17.50*	1a
60	*M = 12.67*	*t(49) = 5.25,* $\mathbf{p < .001}$*,* $d = 0.74$*, 95% CI = [7.82, 17.51]*	1b
120	M = 4.9	$t(49) = 2.23$, $\mathbf{p = .03}$, $d = 0.32$, 95% CI = [0.49, 9.31].	1b
180	M = -6.31	$t(49) = -3.50$, $\mathbf{p = .001}$, $d = -0.49$, 95% CI = [-9.94, −2.69].	1b

Data availability

All data and code is available on OSF.

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, S. E., Matlock, T., & Spivey, M. (2013). Grammatical aspect and temporal distance in motion descriptions. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00337

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

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent Advances in Learning and Motivation* (Vol. 8, pp. 47–89). New York: Academic Press. https://doi.org/10.1016/s0079-7421(08)60452-1.

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

Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *The Journal of Neuroscience, 31*(8), 2801–2814. https://doi.org/10.1523/JNEUROSCI.5003-10.2011

[5] Thanks to reviewer 1, we note that an equally plausible alternative is that any intervening task during the delay, especially one that shifts modality or task demands, may interrupt ongoing extrapolation and thereby reduce representational momentum effect. We leave testing this possibility for future research.

Bergen, B., & Wheeler, K. (2009). Grammatical aspect and mental simulation. *Brain and Language, 112*(3), 150–158. https://doi.org/10.1016/j.bandl.2009.07.002
Bertamini, M. (2002). Representational momentum, internalized dynamics, and perceptual adaptation. *Visual Cognition, 9*(1–2), 195–216. https://doi.org/10.1080/13506280143000395
Bevacqua, L., Loáiciga, S., Rohde, H., & Hardmeier, C. (2021). Event and entity coreference across five languages: Effects of context and referring expression. *Dialogue and Discourse, 12*(2), 192–226.
Binder, C. C. (2022). Time-of-day and day-of-week variations in Amazon Mechanical Turk survey responses. *Journal of Macroeconomics, 71*, Article 103378.
Blalock, L. D. (2013). Mask similarity impacts short-term consolidation in visual working memory. *Psychonomic bulletin & review, 20*(6), 1290–1295. https://doi.org/10.3758/s13423-013-0461-9
Broadbent, D. E., & Broadbent, M. H. (1981). Recency effects in visual memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33A*(1), 1–15. https://doi.org/10.1080/14640748108400762
Carreiras, M., Carriedo, N., Alonso, M. A., & Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Memory & Cognition, 25*(4), 438–446. https://doi.org/10.3758/BF03201120
Casey, L. S., Chandler, J., Levine, A. S., Proctor, A., & Strolovitch, D. Z. (2017). Intertemporal differences among MTurk workers: Time-based sample variations and implications for online data collection. *SAGE Open, 7*(2), Article 2158244017712774.
Comrie, B. (1976). *Aspect (Cambridge Textbooks in Linguistics)*. Cambridge: Cambridge University.
Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X
Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The Demise of Short-Term memory Revisited: Empirical and Computational Investigations of Recency Effects. *Psychological Review, 112*(1), 3–42. https://doi.org/10.1037/0033-295X.112.1.3
Demirdache, H., & Martin, F. (2015). Agent control over non-culminating events. In E. Barrajón, J. L. Cifuentes, & S. Rodríguez (Eds.), *Verb classes and aspect* (pp. 185–217). Amsterdam: John Benjamins. https://doi.org/10.1075/ivitra.9.09dem.
Fallon, J., & Pylkkänen, L. (2024). Language at a glance: How our brains grasp linguistic structure from parallel visual input. *Science advances, 10*(43), Article eadr9951. https://doi.org/10.1126/sciadv.adr9951
Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic bulletin & review, 18*(1), 150–157. https://doi.org/10.3758/s13423-010-0021-5
Feist, M. I., & Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition, 35*(2), 283–296. https://doi.org/10.3758/BF03193449
Ferreira, F., & Barker, M. (2024). Perceptual clauses as units of production in visual descriptions. *Topics in Cognitive Science*.
Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 182–196.
Finke, R. A., & Freyd, J. J. (1985). Transformations of visual memory induced by implied motions of pattern elements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 780–794. https://doi.org/10.1037/0278-7393.11.1-4.780
Finke, R. A., Freyd, J. J., & Shyi, G. C. W. (1986). Implied velocity and acceleration induce transformations of visual memory. *Journal of Experimental Psychology: General, 115*, 175–188.
Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of experimental psychology. General, 148*(7), 1105–1123. https://doi.org/10.1037/xge0000560
Blender Foundation. (2020). Blender (Version 2.82) [Computer software]. https://www.blender.org.
Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics, 33*(6), 575–581. https://doi.org/10.3758/BF03202940
Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review, 94*, 427–438.
Freyd, J. J. (1992). Dynamic representations guiding adaptive behavior. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action, and cognition: Towards bridging the gap* (pp. 309–323). Dordrecht, The Netherlands: Kluwer Academic Publishers.
Freyd, J. J. (1993). Five hunches about perceptual processes and dynamic representations. In D. Meyer, & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 99–119). Cambridge, MA: MIT Press.
Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(1), 126–132. https://doi.org/10.1037/0278-7393.10.1.126
Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition, 83*(1), 49–79.
Gentner, D., & Loftus, E. F. (1979). Integration of verbal and visual information as evidenced by distortions in picture memory. *The American Journal of Psychology, 92*(2), 363–375. https://doi.org/10.2307/1421930
Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making, 26*(3), 213–224.
Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498.
Hafri, A., Boger, T., & Firestone, C. (2022). Melting Ice with your mind: Representational Momentum for Physical States. *Psychological Science, 33*(5), 725–735. https://doi.org/10.1177/09567976211051744
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? *The Journal of Neuroscience, 32*(17), 5795–5803. https://doi.org/10.1523/JNEUROSCI.6294-11.2012
Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review, 12*(5), 822–851. https://doi.org/10.3758/bf03196775
Hubbard, T. L., & Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics, 44*(3), 211–221. https://doi.org/10.3758/BF03206290
Huette, S., Winter, B., Matlock, T., Ardell, D. H., & Spivey, M. (2014). Eye movements during listening reveal spontaneous grammatical processing. *Frontiers in Psychology, 5*, 410. https://doi.org/10.3389/fpsyg.2014.00410
Jackendoff, R. (2023). The parallel architecture in language and elsewhere. *Topics in Cognitive Science, 17*(4), 822–831.
Jardón, N., Marx, E., & Wittenberg, E. (2025). Is there a perfect state? Experimental evidence from English and Spanish for the perfect-as-state hypothesis. Glossa: a journal of general linguistics, 10(1).
Kamp, H., & Reyle, U. (1993). *From discourse to logic: An introduction to model-theoretic semantics of natural language, formal logic, and discourse representation theory*. Dordrecht: Kluwer.
Kang, X., Joergensen, G. H., & Altmann, G. T. M. (2020). The activation of object-state representations during online language comprehension. *Acta Psychologica, 210*, Article 103162. https://doi.org/10.1016/j.actpsy.2020.103162
Kerzel, D. (2000). Eye movements and visible persistence explain the mislocalization of the final position of a moving target. *Vision research, 40*(27), 3703–3715. https://doi.org/10.1016/s0042-6989(00)00226-1
Kerzel, D., Hecht, H., & Kim, N. G. (2001). Time-to-passage judgments on circular trajectories are based on relative optical acceleration. *Perception & Psychophysics, 63*(7), 1153–1170. https://doi.org/10.3758/bf03194531
Lakusta, L., & Carey, S. (2015). Twelve-month-Old infants' Encoding of Goal and Source Paths in Agentive and Non-Agentive Motion events. *Language learning and development : the official journal of the Society for Language Development, 11*(2), 152–157. https://doi.org/10.1080/15475441.2014.896168
Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition, 96*(1), 1–33. https://doi.org/10.1016/j.cognition.2004.03.009
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. https://doi.org/10.1111/j.1551-6709.2011.01220.x
Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning & Verbal Behavior, 13*(5), 585–589. https://doi.org/10.1016/S0022-5371(74)80011-3
Lupyan, G., Abdel Rahman, R., Boroditsky, L., & Clark, A. (2020). Effects of Language on Visual perception. *Trends in cognitive sciences, 24*(11), 930–944. https://doi.org/10.1016/j.tics.2020.08.005
Madden, C. J., & Therriault, D. J. (2009). Verb aspect and perceptual simulations. *The Quarterly Journal of Experimental Psychology, 62*(7), 1294–1303. https://doi.org/10.1080/17470210802696088
Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition, 31*(5), 663–672. https://doi.org/10.3758/BF03196106
Magliano, J. P., & Schleich, M. C. (2000). Verb aspect and situation models. *Discourse Processes, 29*(2), 83–112.
Maier, M., & Abdel Rahman, R. (2019). No matter how: Top-down effects of verbal and semantic category knowledge on early visual perception. *Cognitive, affective & behavioral neuroscience, 19*(4), 859–876. https://doi.org/10.3758/s13415-018-00679-8
Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General, 129*(3), 361–368. https://doi.org/10.1037/0096-3445.129.3.361
Marx, E., & Wittenberg, E. (2025). Temporal construal in sentence comprehension depends on linguistically encoded event structure. *Cognition, 254*, Article 105975. https://doi.org/10.1016/j.cognition.2024.105975
Marx, E., Iwan, O., & Wittenberg, E. (2025). Dynamicity Predicts Inferred Temporal Order in complex sentences: Evidence from English, German, and polish. *Cognitive Science, 49*(2), Article e70038. https://doi.org/10.1111/cogs.70038
Merz, S., Deller, J., Meyerhoff, H. S., Spence, C., & Frings, C. (2019). The contradictory influence of velocity: Representational momentum in the tactile modality. *Journal of Neurophysiology, 121*(6), 2358–2363. https://doi.org/10.1152/jn.00128.2019
Merz, S., Soballa, P., Spence, C., & Frings, C. (2022). The speed prior account: A new theory to explain multiple phenomena regarding dynamic information. *Journal of Experimental Psychology: General, 151*(10), 2418–2436. https://doi.org/10.1037/xge0001212
Minor, S., Mitrofanova, N., & Ramchand, G. (2022). Fine-grained time course of verb aspect processing. *PLoS One1, 17*(2), Article e0264132. https://doi.org/10.1371/journal.pone.0264132
Misersky, J., Slivac, K., Hagoort, P., & Flecken, M. (2021). The state of the onion: Grammatical aspect modulates object representation during event comprehension. *Cognition, 214*, Article 104744. https://doi.org/10.1016/j.cognition.2021.104744
Mittwoch, A. (2008). The English resultative perfect and its relationship to the experiential perfect and the simple past tense. *Linguistics and Philosophy, 31*(3), 323–351.

Moens, M. (1987). *Tense, aspect and temporal reference*. University of Edinburgh. PhD Thesis.
Moens, M., & Steedman, M. (1988). Temporal ontology and temporal reference. *Computational linguistics, 14*(2), 15–28.
Morrow, D. G. (1985). Prepositions and verb aspect in narrative understanding. *Journal of Memory and Language, 24*(4), 390–404. https://doi.org/10.1016/0749-596X(85)90036-1
Morrow, D. G. (1990). Spatial models, prepositions, and verb-aspect markers. *Discourse Processes, 13*(4), 441–469. https://doi.org/10.1080/01638539009544769
Mozuraitis, M., Chambers, C. G., & Daneman, M. (2013). Younger and older adults' use of verb aspect and world knowledge in the online interpretation of discourse. *Discourse Processes, 50*(1), 1–22. https://doi.org/10.1080/0163853X.2012.726184
Müsseler, J., Van Der Heijden, A. H. C., Mahmud, S. H., Deubel, H., & Ertsey, S. (1999). Relative mislocalization of briefly presented stimuli in the retinal periphery. *Perception & Psychophysics, 61*(8), 1646–1661. https://doi.org/10.3758/BF03213124
Papafragou, A. (2010). Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive Science, 34*(6), 1064–1092. https://doi.org/10.1111/j.1551-6709.2010.01107.x
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
Parsons, T. (1990). *Events in the Semantics of English*. MIT Press.
Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006
Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017
Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-01193-y
Prystauka, Y., Hao, J., Cabrera Perez, R., & Rothman, J. (2024). Lexical interference and prediction in sentence processing among russian heritage speakers: An individual differences approach. *Journal of Cultural Cognitive Science. Advance online publication.*. https://doi.org/10.1007/s41809-024-00148-4
Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences, 7*(5), 187–189. https://doi.org/10.1016/S1364-6613(03)00092-5
Reed, C. L., & Vinson, N. G. (1996). Conceptual effects on representational momentum. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 839–850. https://doi.org/10.1037/0096-1523.22.4.839
Regier, T., & Zheng, M. (2007). Attention to endpoints: A cross-linguistic constraint on spatial meaning. *Cognitive Science, 31*, 705–719.
Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2001). Visual short-term memory in the first year of life: Capacity and recency effects. *Developmental Psychology, 37*(4), 539–549. https://doi.org/10.1037/0012-1649.37.4.539
Sakarias, M., & Flecken, M. (2019). Keeping the result in sight and mind: General cognitive principles and language-specific influences in the perception and memory of resultative events. *Cognitive Science, 43*(1), 1–30. https://doi.org/10.1111/cogs.12708
Santin, M., van Hout, A., & Flecken, M. (2021). Event endings in memory and language. *Language, Cognition and Neuroscience, 36*(5), 625–648. https://doi.org/10.1080/23273798.2020.1868542
Smith, C. (1991). *The Parameter of Aspect*. Kluwer Academic Publishers.
Solomon, S. H., Hindy, N. C., Altmann, G. T., & ThompsonSchill, S. L. (2015). Competition between mutually exclusive object states in event comprehension. *Journal of Cognitive Neuroscience., 27*(12), 2324–2338.
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
Truitt, T. P., & Zwaan, R. A. (1997). November). Verb aspect affects the generation of instrumental inferences. *Paper presented at the 38th annual meeting of the Psychonomic Society*.
Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373. https://doi.org/10.1037/h0020071
van Hout, A. (2016). Lexical and Grammatical Aspect. In J. Lidz, W. Snyder, & J. Pater (Eds.), *The Oxford Handbook of Developmental Linguistics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199601264.013.25.
Vinson, N. G., & Reed, C. L. (2002). Sources of object-specific effects in representational momentum. *Visual Cognition, 9*(1–2), 41–65. https://doi.org/10.1080/13506280143000313
Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1436–1451. https://doi.org/10.1037/0096-1523.32.6.1436
von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics, 50*(4), 833–867. https://doi.org/10.1515/ling-2012-0026
Wampler, J., & Wittenberg, E. (2023). Discourse structure affects reference resolution to events in English: Evidence from a new paradigm. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 45, No. 45).
Wang, Y., & Gennari, S. P. (2019). How language and event recall can shape memory for time. *Cognitive Psychology, 108*, 1–21.
Witney, A. G., Vetter, P., & Wolpert, D. M. (2001). The influence of previous experience on predictive motor control. NeuroReport: For Rapid Communication of. *Neuroscience Research, 12*(4), 649–653. https://doi.org/10.1097/00001756-200103260-00007
Wittenberg, E., & Levy, R. (2017). If you want a quick kiss, make it count: How choice of syntactic construction affects event construal. *Journal of Memory and Language, 94*, 254–271. https://doi.org/10.1016/j.jml.2016.12.001
Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology, 11*(18), R729–R732. https://doi.org/10.1016/S0960-9822(01)00432-8
Zhou, P., Crain, S., & Zhan, L. (2014). Grammatical aspect and event recognition in children's online sentence comprehension. *Cognition, 133*(1), 262–276.