Journal of Memory and Language 149 (2026) 104764

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Representing object state change during sentence comprehension: Boundary conditions and language-specific constraints

Channing E. Hambric [a,*], Jon Sides [a], Kevin J. Holmes [b], Abhilasha A. Kumar [a]

[a] *Department of Psychology, Bowdoin College, United States*
[b] *Department of Psychology, Reed College, United States*

ARTICLE INFO

Keywords:
State change
Mental representation
Implicit object properties
Situation model
Language comprehension

ABSTRACT

Language comprehension entails forming contextually appropriate representations of described entities and events. Previous work suggests that non-visual object properties like weight are automatically integrated into event models during comprehension. Horchak and Garrido (2021) found that Portuguese speakers' response times were faster when the state of a presented object (e.g., a smashed tomato) matched the event implied by the preceding sentence (e.g., *You drop a bowling ball on a tomato*). In a pre-registered replication in English (Experiment 1), we failed to replicate this sentence-picture match effect. To explore the reasons for this non-replication, we designed three pre-registered extensions to examine the effects of another type of state change (slicing, Experiment 2), the role of implicit target object properties (squashability, Experiment 3), and the role of syntactic structure (sentence focus, Experiment 4). Contrary to previous work, we observed a match effect on both accuracy and response times in Experiment 4, but the effect on response times was only present when the target object was the focus (i.e., subject) of sentences with a relatively weak causal structure. Follow-up surveys comparing readers' interpretation of the original Portuguese stimuli and their English translations indicated that the syntactic structure of Portuguese amplified the relative salience of the objects described in the state change event. Overall, these findings suggest that the representation of object state changes during language comprehension depends on the interaction of causal structure, object properties, and language-specific syntactic constraints.

Introduction

During language comprehension, readers and listeners use sentential context to form appropriate mental representations of described objects and events (Glenberg et al., 1999; Glenberg et al., 1999; Hoeben Mannaert et al., 2019). They also use pre-existing conceptual knowledge to inform their interpretation of what is described. For example, imagine that you read the sentence *The woman dropped her popsicle on the sidewalk*. What form of popsicle comes to mind? It would likely not be the upright, canonical version, but rather a variant with the popsicle on its side, perhaps even broken into pieces. Now, suppose the sentence were extended to *The woman dropped her popsicle on the sidewalk on a hot summer day*. In this case, given your pre-existing knowledge that popsicles are prone to melting in the heat, you might represent the popsicle in a *melted* state. People thus use a variety of cues to construct "situation models" (Zwaan & Radvansky, 1998) or "mental models" (Johnson-Laird, 1983) of described events during language comprehension (Flusberg et al., 2024). In the present work, we sought to investigate the types of object characteristics that are automatically integrated into these models, how object representations are updated as events unfold, and whether mental simulation processes vary across different syntactic structures.

Simulation and grounded cognition

The richness of mental models has previously been demonstrated via the widely studied action-sentence compatibility effect (ACE). In a typical paradigm investigating the ACE, participants read a sentence describing an action and then press one of two buttons to indicate whether the sentence is sensible or nonsensical (see Winter et al., 2022 for an extensive review). The key finding is that response times are faster when the action described in the sentence matches the direction of the action used by participants to make their response. For example, sentences implying movement away from the body (e.g., *You delivered the*

* Corresponding author at: Department of Psychology, Bowdoin College, 5600 College Station, Brunswick, ME 04011, United States.
E-mail address: c.hambric@bowdoin.edu (C.E. Hambric).

https://doi.org/10.1016/j.jml.2026.104764
Received 11 June 2025; Received in revised form 27 March 2026; Accepted 30 March 2026

pizza to Andy) are verified faster when participants must move their hand away from (versus toward) their body to press the correct response key (Glenberg & Kaschak, 2002).

This finding has been taken as support for embodied theories of language comprehension (Glenberg & Kaschak, 2002; Greco, 2021; Kaschak & Borreggine, 2008; Winter et al., 2022). For example, Barsalou (1999) argued that conceptual representations, including those derived during language comprehension, are directly analogous to and grounded in the same neural systems that support perception. Similarly, Glenberg et al. (1999) suggested that perceptual representations—not abstract concepts or schemas—constitute the background knowledge used to construct event models during language comprehension. In this view, comprehending a sentence describing a particular motor movement (e.g., handing a pizza to Andy) entails activating the same neural circuits actually used to enact the movement, thus priming the same motor action when it is used to give a response (Glenberg & Kaschak, 2002).

Although theoretically compelling, these claims have recently been called into question due to replication failures (Papesh, 2015), including large, pre-registered, multilab attempts to replicate the ACE (Morey et al., 2022). In addition, a recent meta-analysis by Winter et al. (2022) indicated that a number of factors can affect the robustness of the observed ACE, including stimulus selection and the delay between the sentence and response. However, despite these caveats, Winter et al. (2022) observed a small but significant ACE across studies, consistent with the construction of action-based event models during language comprehension.

Mental models of object state change

The ACE has also inspired a body of work probing the simulation of objects, not just motor movements, in mental models of events during language comprehension. According to an embodied perspective, linguistic input is indexed to perceptual symbols to derive object-specific affordances, and these affordances are meshed under perceptual and linguistic constraints to create appropriate simulations of comprehended events (Glenberg et al., 1999; Glenberg et al., 1999; Kaschak & Glenberg, 2000; Stanfield & Zwaan, 2001). The current body of evidence indicates that readers readily integrate surface-level object characteristics, such as implied spatial orientation, into mental models of events.

The sentence-picture verification task (e.g., Zwaan & Pecher, 2012) has been used to probe how objects are mentally represented during comprehension. In a typical version of this paradigm, participants read a sentence describing a particular event and are then presented with an image of an object. The task is to determine whether the pictured object was mentioned in the preceding sentence, and the response time to make this judgment is recorded. Early work by Stanfield and Zwaan (2001) asked participants to read sentences like *The carpenter pounded the nail into the wall* and *The carpenter pounded the nail into the floor*. Participants were faster to verify the pictured object when it was shown in the orientation implied by the sentence (e.g., a horizontal vs. vertical nail). This paradigm has also been extended to investigate the representation of other explicit object properties, such as object shape and color (Connell, 2005; Connell, 2007; Stanfield & Zwaan, 2001; Zwaan et al., 2002). For example, across a series of replications, Zwaan and Pecher (2012) found significant match effects for object orientation, shape, and color, the latter of which contradicts the color *mis*match effect found by Connell (2005) and Connell (2007).[1] Nevertheless, this work provided evidence that the explicit properties of orientation and shape are automatically integrated into event models during comprehension.

More recently, Kang et al. (2020a) manipulated the *state* of the object rather than its orientation or shape. They asked participants to read sentences that described either substantial change to an object's state (e.g., *The woman dropped the trash can*) or minimal to no change (e.g., *The woman placed the trash can*). Participants were faster to verify objects pictured in a non-canonical, modified state (e.g., a tipped-over trash can) when they followed sentences implying a substantial change. Conversely, participants were faster to verify objects pictured in a canonical state following sentences implying little to no change. These findings suggest that readers infer information about object state from linguistic input and incorporate this information into their event models. Other methodologies have also been developed to assess mental representations of objects during sentence comprehension. For example, Kang et al. (2020b) used the Visual World Paradigm to assess gaze shifts between described objects (see also Krass et al., 2025). In this paradigm, a sentence is auditorily presented while participants view pictures depicting two versions of a target object and two unrelated distractors. When participants heard sentences describing state change events (e.g., *The chef will chop an onion*), they fixated longer on the object in the state that matched the language context (e.g., a chopped onion) than the same object in an incongruent state (e.g., a whole onion).

There is also evidence that readers incorporate *implicit* object properties into event models during comprehension (Bransford et al., 1972; Briner et al., 2014; Sachs, 1967). In a recent series of experiments, Horchak and Garrido (2021) investigated whether the implicit property of weight is integrated into mental models. They also examined whether people maintain access to both canonical and modified object states after reading a sentence implying substantial change to an object. The latter question stems from the proposal that event representations are comprised of collections of intersecting object histories (Altmann & Ekves, 2019). According to this account, people encode and maintain access to multiple distinct states of an object over the course of an event. Building on this idea, Horchak and Garrido (2021) argued that the modified state of an object resulting from an external action cannot be wholly divorced from its initial state. That is, to know what a *squashed tomato* looks like, one must also picture what it looked like *prior* to being squashed. Thus, according to Horchak and Garrido, reading a sentence that implies substantial change to an object's state does not replace the canonical representation of the object with a modified version. Instead, readers represent multiple, competing object states during comprehension (also see Hindy et al., 2012; Solomon et al., 2015).

In Horchak and Garrido's (2021) study, participants read sentences in which a heavy or light object (e.g., a bowling ball or a balloon) was dropped onto a malleable object (e.g., a tomato). Critically, the weight of the dropped object was not explicitly expressed in the text and thus could only be inferred from pre-existing knowledge. After reading each sentence, participants were presented with a picture of either an intact, canonical version of the malleable object or a smashed, non-canonical version. The results showed that participants were faster to verify objects in a non-canonical state after reading a "heavy" sentence (i.e., when a *bowling ball* was the acting object) compared to a "light" sentence (i.e., when a *balloon* was the acting object). This sentence-picture match effect suggests that participants mentally "squashed" the target object after integrating the dropped object's weight into their event model. Importantly, in contrast to the work by Kang et al. (2020a), which implied object state change by manipulating the verb phrase (e.g., *dropped* vs. *placed*), items in a canonical state were *not* verified faster after light sentences (which imply little to no state change). This lack of a cross-over interaction suggests that sentences that imply state change through implicit world knowledge while keeping the action constant do not elicit the same type of mental simulation as those that describe state change explicitly.

Horchak and Garrido (2021) offered two possible interpretations of their findings: either only the end-state of the object (the "smashed" state) is encoded in the mental model (Zwaan & Pecher, 2012), or both the original and modified states are simultaneously active (Altmann & Ekves, 2019; Hindy et al., 2012; Solomon et al., 2015). They argued that

[1] We also conducted a study examining the interaction between object color and state, though we limit the discussion of this experiment to the supplemental materials.

the latter possibility was supported by the finding that canonical objects were verified no faster after a light sentence than a heavy sentence, suggesting that both the original and modified states were accessible even when the preceding sentence implied substantial state change. The authors concluded that not only are implicit object properties integrated into event models, but readers appear to maintain access to the varying object states implied by the sentence narrative (Altmann & Ekves, 2019). In subsequent experiments, Horchak and Garrido (2021) replicated the sentence-picture match effect for non-canonical objects, and the lack thereof for canonical objects, across several fine-grained manipulations, including different acting objects (e.g., *brick* vs. *sponge*), different tenses (*You drop* vs. *You dropped*), and different phrasal structures (*You dropped a bowling ball on a tomato* vs. *A brick fell on a tomato*). In Experiment 1 of the current work, we sought to replicate this key finding in an English-speaking sample.

Extending the scope of object state change

Although Horchak and Garrido's (2021) findings shed light on the representation of implicit object properties, their work—and most other research on object state change—has focused solely on comparing intact objects to objects in a state of destruction (tipped over, smashed/squashed; see Kang et al., 2020a; Horchak & Garrido, 2024). However, the state of an object may be modified in a variety of other ways. For example, in addition to an action causing an object to be *squashed* or deformed, slicing an object changes its state from intact to *sliced*. While some past work has included slicing (or chopping) as one form of state change within a broader class of "modified" representations (e.g., Kang et al., 2020b), to date there has been little work isolating this specific type of state change. It thus remains unknown if readers simultaneously represent the whole and part states of an object (e.g., an intact and a sliced tomato) during comprehension or if the intact representation of the object is entirely replaced by the part state when the act of *slicing* unfolds within the context of the sentence (Altmann & Ekves, 2019). We explore this question in Experiment 2 of the present work.

There is some evidence suggesting that the representation of an object's final state depends on whether the action has already been completed. Across experiments, Kang et al. (2020a) manipulated whether the sentences described state change using past tense (*The woman dropped the ice cream,* Experiment 2) or future tense (*The woman will drop the ice cream,* Experiment 3). Whereas a significant match effect was observed for non-canonical objects following sentences in either tense, canonical objects were verified faster after neutral sentences compared to sentences implying state change, but only in the past tense condition. This suggests that readers use linguistic input to form a contextually relevant iteration of an object, but that the grammatical structure of the sentence plays a key moderating role.

Importantly, the majority of past work using the sentence-picture verification paradigm has not manipulated the implicit characteristics of the target objects themselves (Horchak & Garrido, 2021). This issue was partially addressed by Kang et al. (2020b) in an experiment where target objects had different degrees of intrinsic squashability (e.g., an egg vs. a penny). However, the malleability of the target object was varied in order to manipulate the degree of state change, rather than to assess the target object's *implicit* properties. Within the Visual World Paradigm, participants who heard sentences describing substantial state change (e.g., *The girl will stomp on the egg*) fixated longer on objects in "situationally appropriate" states (e.g., a smashed egg) than objects in an incongruent state (e.g., a whole egg), but that the latter objects nevertheless garnered attention. However, it remains unknown how the implicit properties of target objects interact with the implicit properties of other entities in the described event, and whether simulation processes are themselves *constrained* by these intrinsic properties. We address these open questions in Experiment 3.

Lastly, beyond tense, only limited research has examined how sentence structure may influence the representation of object state change. Horchak and Garrido (2021) observed a match effect using sentences with an explicit, intentional agent enacting change upon the object (e.g., *You drop a bowling ball on a tomato*) as well as sentences with no agent mentioned (e.g., *A brick fell on a tomato*). However, differing sentence structures can influence simulation during narrative comprehension (Jin et al., 2024; Lee & Kaiser, 2021; Sato et al., 2013). This issue may be especially important to consider when comparing sentences with different causal structures and complexity. The sentence structures used by Horchak and Garrido (2021) implied a direct causal link between the acting object (e.g., the bowling ball) and the target object (e.g., the tomato), but the strength of the sentence-picture match effect may be moderated by the strength of this causal link. Specifically, clear causality may elicit more robust event models with more salient representations of state change. However, other work (e.g., Mason & Just, 2007) suggests that in the case of ambiguous sentences, multiple interpretations are salient during comprehension. This raises the intriguing possibility that sentence-picture match effects may be stronger for sentences with underspecified causal structures, which invite comprehenders to actively work out the relationship between the acting and target objects.

The moderating role of sentence structure is also crucial when assessing mental simulation in languages with different syntactic constraints. For example, in null subject languages (such as European Portuguese, used in Horchak & Garrido, 2021), the translated equivalent of *You drop a bowling ball on a tomato* begins with the conjugated form of the verb *deixar* (to let [drop]) rather than the explicit subject pronoun "You" required by English grammar. Presenting an explicit subject prior to the verb may influence what readers consider to be the focal noun phrase of the sentence (Alonso-Ovalle et al., 2002; del Río et al., 2011; Santos, 2005). Consistent with this possibility, linguists have proposed that speakers of languages in which both null and overt subject structures are acceptable (including Brazilian Portuguese, Italian, and Spanish) use overt subjects to disambiguate antecedents and shift attention to a specific referent (Frascarelli, 2007; Serratrice, 2008; Soares et al., 2019). This may be a key factor when evaluating the current body of evidence, which includes languages that utilize a variety of syntactic structures. Such differences, though seemingly superficial, may modulate the salience of object representations during sentence comprehension. In Experiment 4, we explore the intertwined issues of sentence structure and language-specific syntactic constraints.

The present studies

We conducted four pre-registered experiments investigating the representation of object state change during written sentence comprehension. Experiment 1 was a replication of Horchak and Garrido's (2021) Experiment 1, with the exception that the original study was conducted in Portuguese and the present work was conducted in English. To preview the results, we failed to replicate the sentence-picture match effect in reaction times observed for non-canonical objects after reading heavy sentences. In follow-up experiments, we investigated a different type of state change and whether implicit properties of target objects influence the match effect. Specifically, Experiment 2 explored the representation of state change implied by sentences in which an object is *sliced*, and Experiment 3 manipulated the implicit *squashability* of the target object. Both experiments failed to find a significant match effect. In Experiment 4a, we investigated a potential explanation for the replication failures: cross-language differences in syntactic structure, which may influence sentential focus and subsequent mental representations of object state changes. To test this possibility, we manipulated the sentence structure to place the malleable target object as either the focal subject or the direct object of the sentence. We found that pictures in a non-canonical state were verified significantly faster following heavy than light sentences, but only when the target object was the focus of the sentence. However, direct comparison to Experiment 1 was complicated by the use of sentence structures with different degrees of implied causality. We thus conducted a follow-up study (Experiment 4b) using sentences that more closely approximated the stimuli used in Experiment 1, but this time we failed to replicate the moderating influence of sentence focus. To clarify these results, we conducted a follow-up survey with English and Portuguese speakers (Appendix B) and a separate survey with English speakers (Appendix C) to provide insight into how readers implicitly interpret the various syntactic structures used across experiments.

Together, our work suggests that the representation of object state change during sentence comprehension depends on multiple situational and linguistic constraints, including sentence focus, causal structure, object affordances, and readers' implicit interpretations of syntax. By systematically varying these factors across experiments, the present work demonstrates that object state change is more context-sensitive—and theoretically more complex—than prior accounts have suggested.

Experiment 1: English replication

We first conducted a direct English replication of Horchak and Garrido's (2021) Experiment 1. Our replication investigated whether the implicit object property of weight is integrated into mental models during on-line sentence comprehension. Following Horchak and Garrido (2021, Experiment 1; henceforth, the *original study*), we predicted that objects in non-canonical states would be verified faster following sentences that describe dropping a heavy object (i.e., a *bowling ball*) onto them compared to those that describe dropping a light object (i.e., a *balloon*). We predicted no effect of sentence type when verifying objects in (intact) canonical states. The replication was pre-registered (https://aspredicted.org/yr4sm.pdf). Unless otherwise noted, our procedure was identical to the original work.

Method

Participants

We relied on the power analysis conducted by Horchak and Garrido (2021) to set a minimum target sample size of 77 participants. To account for exclusions, we recruited 99 participants from Prolific. Recruitment was restricted to native English speakers. Participant demographics across all experiments can be seen in Table 1. There were no exclusions based on comprehension performance (see below), resulting in a final sample of 99 participants for the accuracy analysis. Following Horchak and Garrido (2021), participants with less than 80% accuracy on target trials were excluded from response time analyses ($n = 6$), resulting in a final sample of 93 participants. Participants with excessive error rates were also excluded from the accuracy analyses in Horchak and Garrido (2021), but we chose to include them in the present accuracy analysis to better capture the full variability of the data.

Design and materials

The experiment was programmed in jsPsych (de Leeuw, 2015) and conducted via *Cognition* (https://www.cognition.run/). We reused all stimuli from Horchak and Garrido's (2021) Experiment 1, with the provided English-translated sentence stimuli in place of the original Portuguese stimuli. The sentence-picture verification task contained 72 trials, with 24 experimental trials and 48 fillers. The filler sentences described seeing an event (e.g., *You see how a donkey is drinking water*) or dropping a different object than in the experimental sentences (e.g., *You drop a cushion on a magazine*). All participants saw the same filler trials in randomized order. Of the experimental trials, 12 sentences were "heavy" (*You drop a bowling ball*…), and the other 12 were "light" (*You drop a balloon*…). Each sentence had a different target object that was easily squashable (e.g., *tomato, blueberry, glasses*, etc.). On these trials, the target objects in the images were either intact (canonical) or smashed (non-canonical). Thus, the sentence-picture pairs fit into one of four *weight* × *object state* conditions: heavy–canonical, heavy–non-canonical, light–canonical, or light–non-canonical (see Fig. 1), with six trials in each condition. Four randomly assigned lists were used to counterbalance the condition in which each object was presented across participants. The corresponding target object (in either a canonical or non-canonical state) was presented after each experimental sentence, such that "yes" (the object was mentioned in the previous sentence) was the correct response for all experimental trials. Half of the filler trials were presented with the object mentioned in the sentence, resulting in 12 sentences in which "yes" was the correct response and 12 in which "no" was the correct response.

Finally, 24 comprehension questions were randomly dispersed throughout the main task to ensure participants fully read and understood the sentences. Comprehension questions were only presented after filler trials. Based on a pilot experiment, we altered the format and wording of the comprehension questions to make them easier to understand in English. For example, the question "*In the sentence a giraffe was playing with a zebra?*" (the English translation provided by Horchak & Garrido, 2021) was changed to "*Was a giraffe mentioned in the previous sentence?*" In addition, the screen was modified to have a black background with blue text to more clearly differentiate these trials from the main task.

Procedure

Participants completed the experiment online. The task began with six practice trials (with feedback), followed by a 72-trial main task with randomly dispersed comprehension questions. In each trial of the sentence-picture verification task (see Fig. 1), participants were presented with a sentence followed by a picture of an object and asked to indicate if the object was mentioned in the preceding sentence. A fixation cross appeared for 1000 ms before the sentence stimulus. Participants pressed the space bar after fully understanding the sentence, prompting a fixation cross to appear for 500 ms followed by the target image. Upon seeing the image, participants responded with the 'S' key for yes and the 'N' key for no, which mapped onto the onsets of yes and no in the original Portuguese study. Response accuracy and response time from picture onset were recorded. Following the main task, participants completed a demographic questionnaire.

Results

Data treatment

We conducted all data processing and analyses in R Studio version 4.2.0 (R Core Team, 2024). Before analyses, filler and comprehension trials were removed from the data. All participants correctly answered more than 50% of the comprehension questions, resulting in no comprehension-based exclusions. For response time (RT) analyses, we removed incorrect ('no') responses and excluded participants who responded correctly on less than 80% of experimental trials ($n = 6$). Following Horchak and Garrido (2021), a log-10 transformation was used to normalize RT data, and RTs that were more than three median absolute deviations away from their condition median were removed.

Accuracy analysis

Overall task accuracy was 95.24%. Using the *lme4* package (Bates et al., 2015) and treatment contrast coding, we fit a maximal logit mixed-effects model using weight, object state, and their interaction as fixed effects. The random effects structure included random intercepts for participants and items and by-participant random slopes for weight, object state, and the interaction term.[2] As shown in Fig. 2 (left panel), there were no main effects of weight ($b = 0.89$, $z = 1.09$, $p = .277$) or object state ($b = 0.10$, $z = 0.09$, $p = .931$), and the interaction between weight and object state was not significant ($b = 0.26$, $z = 0.19$, $p = .849$). This finding is consistent with Horchak and Garrido (2021), who did not

[2] While Horchak and Garrido (2021) initially fit a model with the same structure, they removed the nonsignificant interaction term for the primary analyses.

Table 1
Demographics.

Variable	Experiment 1	Experiment 2	Experiment 3	Experiment 4a	Experiment 4b
N	93	89	237	186	157
Gender					
Male	48%	56%	56%	57%	56%
Female	52%	44%	43%	42%	43%
Other	0%	0%	1%	1%	1%
Mean Age (SD)	35.6 (13.9)	33.9 (12.5)	33.1 (13.7)	36.2 (14.8)	44.3 (12.4)
Race					
Asian	8%	8%	5%	5%	8%
Black	14%	12%	13%	20%	13%
Latine	2%	1%	4%	8%	10%
Multiracial	5%	6%	9%	5%	6%
White	70%	65%	68%	64%	71%
Other Race	3%	2%	1%	5%	0%

Demographics. Final sample size after exclusions (*N*), mean age, gender, and race/ethnicity of participants across all experiments.

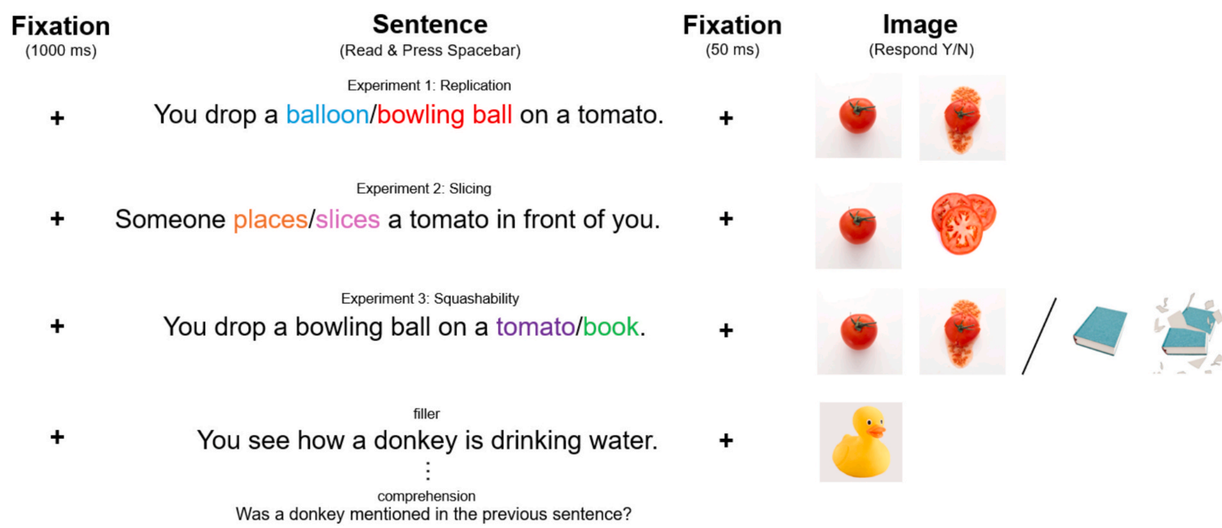

Fig. 1. Overview of Procedures for Experiments 1–3. Conditions and sample stimuli from Experiments 1–3. Experiment 1 was an English replication of Experiment 1 in Horchak and Garrido (2021), Experiment 2 used a slicing verb to imply state change, and Experiment 3 manipulated the squashability of the target object. Filler items were included in all experiments, but comprehension questions were only included in Experiment 1.

observe any effects on accuracy measures.

Response time analysis

Following Horchak and Garrido (2021), we fit a linear mixed-effects model to the transformed RT data with the same fixed and random effects structure as the accuracy model. In contrast to the findings reported by Horchak and Garrido (2021), as shown in Fig. 2 (right panel), there were no main effects of weight ($b = .0008$, $t = 0.10$, $p = .920$) or object state ($b = .03$, $t = 1.53$, $p = .131$), and the interaction between weight and object state was not significant ($b = .004$, $t = -0.38$, $p = .704$). Given this null interaction, we did not conduct any follow-up analyses.

Discussion

We sought to replicate evidence that the weight of an acting object is automatically integrated into mental models of events and that readers maintain access to both original and modified object states. Specifically, we attempted to conduct a replication of Experiment 1 from Horchak and Garrido (2021) using the same procedure and translated materials. We failed to replicate the finding that objects in a non-canonical state are verified faster after sentences implying a state change ("heavy" sentences) than not ("light" sentences). This was surprising given the internal replication across experiments in Horchak and Garrido (2021), but is in line with recent failures to replicate the ACE (Morey et al., 2022; Papesh, 2015). There were few differences across experiments, making it unlikely that the failed replication was due to differences in task demands. That said, one critical difference between the two experiments was language: Horchak and Garrido's (2021) study was conducted in (European) Portuguese, while ours was conducted in English. We return to this issue in Experiment 4.

Another possibility is that the present null findings were due to differences in the familiarity of the pictured objects across participant populations. However, given that the assignment of objects to weight-state conditions was fully counterbalanced across experimental versions and our statistical models accounted for item-level variation, this explanation seems unlikely. We also conducted a follow-up norming experiment to measure how well the non-canonical forms of objects corresponded to what participants mentally simulated when comprehending heavy sentences (see Appendix C). Including these normative ratings as covariates did not alter the null results in the primary analysis of reaction times.

The results of Experiment 1 suggest that the match effect may be less robust than Horchak and Garrido's (2021) findings would imply. In the following experiments, we designed several experimental variations with the goal of exploring the boundary conditions that may modulate the match effect. Specifically, we focused on the nature of the state change (Experiment 2), the properties of the target object (Experiment 3), and which objects are the primary focus within a given sentence framing (Experiment 4).

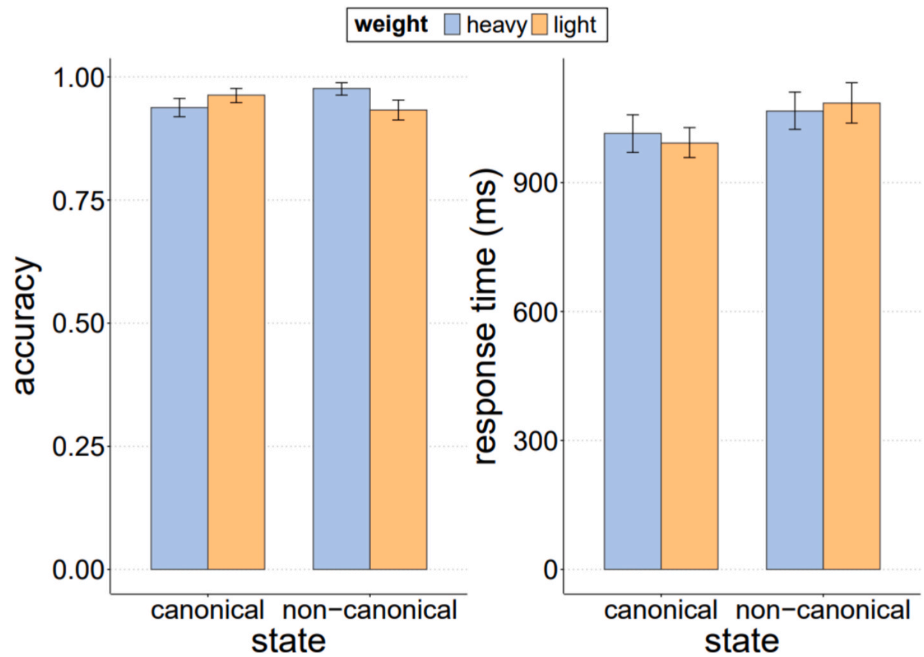

Fig. 2. Experiment 1 Results. Raw accuracy and response times in Experiment 1. Unlike in Experiment 1 in Horchak and Garrido (2021), the weight × object state interaction was not significant in the response time data. Error bars represent bootstrapped 95% confidence intervals.

Experiment 2: Slicing

Experiment 1 failed to replicate the sentence-picture match effect observed by Horchak and Garrido (2021). One potential explanation is that participants had a difficult time appropriately simulating *dropping a bowling ball.* That is, a bowling ball is encountered relatively infrequently, and participants may have never encountered a real-life situation in which a bowling ball was dropped onto another object. Experiment 2 sought to clarify how people simulate other, more common actions during sentence comprehension. Specifically, this experiment used the act of *slicing* to imply that the target object underwent a state change. Since this action is encountered frequently, the resulting state change is highly salient. Moreover, prior work by Kang et al. (2020a) similarly manipulated the verb phrase to imply state change. Therefore, we predicted a similar sentence-picture match effect. Specifically, we predicted that pictures of sliced objects would be verified faster after sentences that described *slicing* the object compared to neutral sentences. The experiment was pre-registered (https://aspredicted.org/gz522.pdf).

Method

Participants

Given the similar design, we used the same power estimation as in the previous experiment and the original study. 106 participants were recruited from Prolific ($N = 100$) and from an introductory psychology course at Bowdoin College ($N = 6$).[3] Six were excluded from all analyses for indicating that English was not their first language. 17 participants were excluded from RT analyses due to low task accuracy, leaving a final sample size of 89 for the RT analysis.

Design, materials, and procedure

The experimental design was similar to the first experiment. Critically, rather than using the weight of a falling object to imply state change, the verb itself was modified to do so (see Fig. 1). Specifically, the verb was modified to imply that the target object was either *sliced* or unaffected. For example, *You drop a bowling ball on…* was modified to:

(1) Someone *slices* an apple in front of you. **(sliced)**
(2) Someone *places* an apple in front of you. **(neutral)**

Thus, the factor *sentence type* refers to the specific verb used in the sentence. Similar to the previous experiment, the subsequent picture was either a canonical, intact version of the target object or a non-canonical (*sliced*) version (see Fig. 1). All objects were commonly sliced foods (e.g., *apple, pizza, tomato*), and all pictures were obtained from Adobe Stock Images. Filler sentences were slightly simplified (e.g., *You look at a [waffle]).* Given that no participants failed the comprehension attention check in Experiment 1, comprehension questions were not used in this experiment. The by-trial procedure was the same as in Experiment 1.

Results

Accuracy analysis

Overall task accuracy was 91.50%. The modeling approach followed Experiment 1. We fit a logit mixed-effects model with sentence type, object state, and their interaction as fixed effects, random intercepts for both participants and items, and a random slope for the main effects and interaction of sentence type and object state by participants. As shown in Fig. 3, there was no significant main effect of sentence type ($b = -2.13$, $z = -1.31, p = .191$) or object state ($b = -2.93, z = -1.84, p = .066$), nor was there a significant interaction ($b = 2.93$, $z = 1.57$, $p = .116$).

Response time analysis

We fit a linear mixed-effects model with the same fixed and random effects structure as the accuracy analysis. There was no main effect of sentence type ($b = 0.01, t = 1.62, p = .107$) or object state ($b = 0.02$, $t = 1.35$, $p = .182$). Critically, there was a significant interaction between

[3] Sample sizes were uneven across recruitment populations, making analysis by data source inappropriate due to low power. In experiments recruiting from both sources (Experiments 2-4a), we confirmed that excluding participants from the smaller population (Bowdoin College) did not significantly change the reported pattern of results.

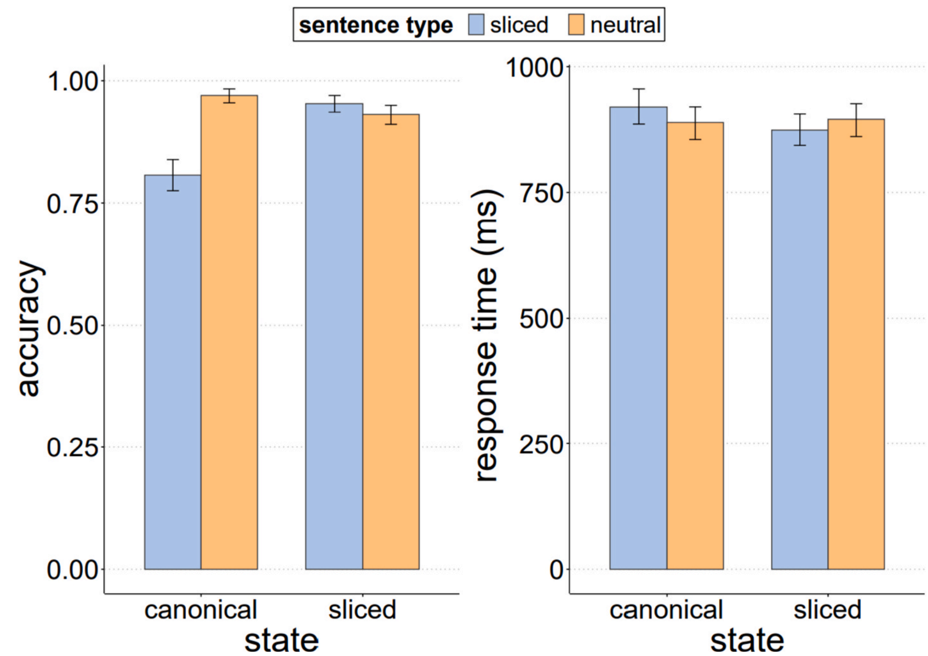

Fig. 3. Experiment 2 Results. Raw accuracy and response times as a function of sentence type and object state in Experiment 2. Error bars represent bootstrapped 95% confidence intervals.

sentence type and object state ($b = -0.02$, $t = -2.43$, $p = .016$). However, follow-up contrasts with Tukey corrections indicated that responses to images of sliced objects were *not* significantly facilitated following *slicing* sentences ($b = 0.01$, $t = -1.61$, $p = .110$), nor were objects in a canonical state verified significantly faster after neutral sentences ($b = 0.01$, $t = 1.64$, $p = .104$), although both effects trended as such (see Fig. 3).

Discussion

The present experiment examined how *slicing* is simulated during sentence comprehension. Sentences either implied state change (via *slicing)* or were neutral, and pictures depicted objects in either a canonical or *sliced* state. We once again failed to observe a significant sentence-picture match effect, even when using the common state change event of *slicing*. Given that Kang et al. (2020a) observed a significant sentence-picture match effect when using a similar verb phrase manipulation and English-speaking sample, this was surprising. One possible explanation is that the high frequency of encountering the objects in their sliced state automatically brought to mind their sliced iteration, regardless of the provided sentential context. The simultaneous representation of both states may have precluded any match effect, similar to the interpretation regarding canonical objects after heavy sentences in Horchak and Garrido (2021). To investigate this idea, in our follow-up norming experiment, we quantitatively measured how well the whole versus sliced versions of objects matched what participants imagined when comprehending a neutral sentence (see Appendix C). Overall, the whole versions of objects were judged to be far more canonical, and the whole and sliced versions were only rated as equally canonical for a single object (*cheesecake*). Moreover, including these normative ratings as covariates did not alter the results of the present analyses. That said, it is still possible that both states of the objects were simulated during real-time comprehension.

It is also notable that the raw accuracy values present a slightly different story than the outputs from the statistical models. Fig. 3 (left panel) appears to show an interaction between sentence type and state, but this interaction was not statistically significant. This discrepancy is due to the large degree of variance captured by the model that included a random slope for the interaction term within participants, which provided a significantly better fit than a simpler model, $\chi^2(9) = 61.36$, $p < .001$. Further examination of the intra-class correlation coefficients (ICC) indicated that a large portion of the total variance was captured by the random effects structure ($ICC = 0.671$), suggesting that individual differences contributed to the significant variance in the overall patterns observed in Experiment 2.

Experiment 3: Squashability

The goal of Experiment 3 was to clarify the null findings in Experiments 1 and 2 by probing another aspect of the sentence-picture match effect: the implicit properties of the pictured target objects. Recall that Kang et al. (2020b) manipulated the inherent squashability of target objects (e.g., a penny vs. an egg) and found that after listening to sentences describing a malleable object changing state (e.g., *The girl will stomp on the egg*), participants fixated longer on objects in a modified state (e.g., a smashed egg) than the same objects in an intact state (e.g., a whole egg). However, the authors only manipulated the state of the "squashable" object (the egg); the unsquashable object (the penny) only ever appeared in a canonical, intact state. In addition, there was only one stimulus object in each condition (e.g., the penny or the egg), leaving it unclear whether this phenomenon extends to a wider array of simulated objects.

We sought to extend this work in three ways. First, whereas Kang et al. (2020b) used eye-tracking within the Visual World Paradigm, we examined whether a similar effect manifests in the context of the sentence-picture verification paradigm, using a more systematic squashability manipulation across a wider variety of objects. Second, related to this issue, it is unclear if mental simulation processes are themselves constrained by properties of the target objects. For example, what would happen if participants comprehended a sentence describing a forceful action enacted on a rigid, unsquashable object (e.g., an eraser) but were then presented with the same object in an unexpectedly altered state (e.g., a squashed/broken eraser)? A match effect might not be observed if the implausible outcome is not simulated. Alternatively, simulation processes might occur automatically during initial comprehension, yielding a match effect regardless of the specific direct object.

Lastly, in contrast to the weight manipulation used by Horchak and Garrido (2021) and in our Experiment 1, Kang et al. (2020b) instantiated state change using a direct causal verb (e.g., "stomped") rather than via implicit knowledge about object properties. Thus, we also sought to assess how implicit properties of the acting and target object interact during comprehension.

To address these open questions, in Experiment 3, all sentences described dropping a heavy *bowling ball* to imply state change, but the implicit *squashability* of the target object (e.g., a *tomato* vs. an *eraser*) varied across items. Critically, we also manipulated object state within items, such that participants verified objects in both an intact, canonical state and a modified, non-canonical state regardless of intrinsic squashability. We predicted that responses to pictures of objects in a squashed (non-canonical) state would be facilitated by sentences implying state change, but only if the target object was intrinsically *squashable.* That is, given that the heavy acting object would only cause state change for squashable objects, this should be reflected in participants' event models. We again predicted no differences for intact (canonical) objects. The experiment was pre-registered (https://aspredicted.org/cn74x.pdf).

Method

Participants

Given the change in design, we conducted a simulation-based power analysis in R using the *simr* package (Green et al., 2016). The simulations indicated that a sample size of approximately 225 was required to achieve 80% power. To account for future exclusions, we recruited 275 participants from an introductory psychology course at Bowdoin College ($N = 57$) and from Prolific ($N = 218$). 17 participants were excluded for indicating that English was not their first language, and 21 participants were excluded for having an overall task accuracy of less than 80%, leaving a final sample of 237 participants for the RT analysis.

Design, materials, and procedure

The experiment retained the picture-state manipulation from the previous two experiments but manipulated the *squashability* of the target object instead of the weight of the falling object. Hence, every sentence described a *bowling ball* being dropped onto a target object, and the target object was either intrinsically *squashable* or *unsquashable.* For example:

(1) You drop a bowling ball onto a *tomato*. **(squashable)**
(2) You drop a bowling ball onto a *book*. **(unsquashable)**

This resulted in a 2 × 2 within-subjects design that manipulated both target object *squashability* (squashable/unsquashable) and *object state* (canonical/non-canonical). Picture stimuli for squashable objects were reused from the original experiment, and unsquashable object pictures were obtained from Adobe Stock Images. Smashed versions of such images were created using Adobe Photoshop (see Fig. 1). New sentences for these unsquashable objects followed the same format as the other sentences. A norming study was conducted to validate the squashability classification for each target object. As can be seen in Fig. 4, objects in the squashable condition were indeed rated as being significantly more squashable than objects in both the filler condition ($b = .59$, $z = 5.30$, $p < .001$) and the unsquashable condition ($b = 1.82$, $z = 15.28$, $p < .001$; see Appendix A for additional details).

As in the previous experiment, comprehension questions were not included in the task. The procedure was identical to previous experiments, and data were processed using the same approach.

Results

Accuracy analysis

Overall task accuracy was 94.30%. We fit a logit mixed-effects model with squashability, object state, and their interaction as fixed effects, random intercepts for both participants and items, and a random slope for the main effects and interaction of squashability and object state by participants. As shown in Fig. 5 (left panel), there was no significant main effect of squashability ($b = 0.68$, $z = 0.80$, $p = .422$) or object state ($b = 1.16$, $z = 1.02$, $p = .305$), nor was there a significant interaction ($b = 1.44$, $z = 0.91$, $p = .361$).

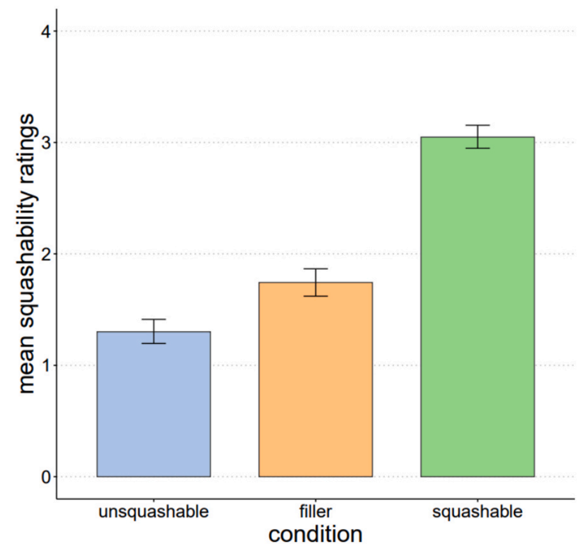

Fig. 4. Squashability Ratings. Squashability ratings (0–4) for Experiment 3 materials, where increasing values indicate greater perceived squashability. Objects in the squashable condition were judged to be significantly more squashable than filler objects and objects in the unsquashable condition. Details of the norming study are available in Appendix A. Error bars represent bootstrapped 95% confidence intervals.

Response time analysis

We fit a linear mixed-effects model with the same fixed and random effects structure as the accuracy analysis, but the model failed to converge. We thus simplified the model to only include random slopes for the main effects of squashability and object state by participants. As shown in Fig. 5 (right panel), there was no significant main effect of squashability ($b = 0.68$, $z = 0.80$, $p = .422$) or object state ($b = 1.16$, $z = 1.02$, $p = .305$), nor was there a significant interaction ($b = 1.44$, $z = 0.91$, $p = .361$). Thus, as in the previous two experiments, even when manipulating the implicit properties of the target objects themselves, we failed to observe a significant sentence-picture match effect.

Discussion

We sought to examine whether implicit target object properties, such as *squashability,* are represented in mental models of state change during comprehension. Specifically, objects were inherently squashable (e.g., *tomato*) or unsquashable (e.g., *book*), and pictures depicted the objects in a canonical or non-canonical (squashed) state. We failed to find a significant match effect, such that non-canonical versions of squashable objects were not verified significantly faster than those of unsquashable objects. Notably, this finding extends the work of Kang et al. (2020b), suggesting that match effects driven by implicit object properties only arise if the altered state is actually viable for a given object.

To probe these null results, we conducted additional analyses using the same model structures except that we included the normative squashability rating for each item (obtained via the norming study described in Appendix A) as a continuous predictor of squashability (rather than the binary un/squashable classification). In the accuracy analysis, there was a significant interaction between state and squashability ($b = -1.15$, $t = -2.10$, $p = .035$), where, surprisingly, accuracy

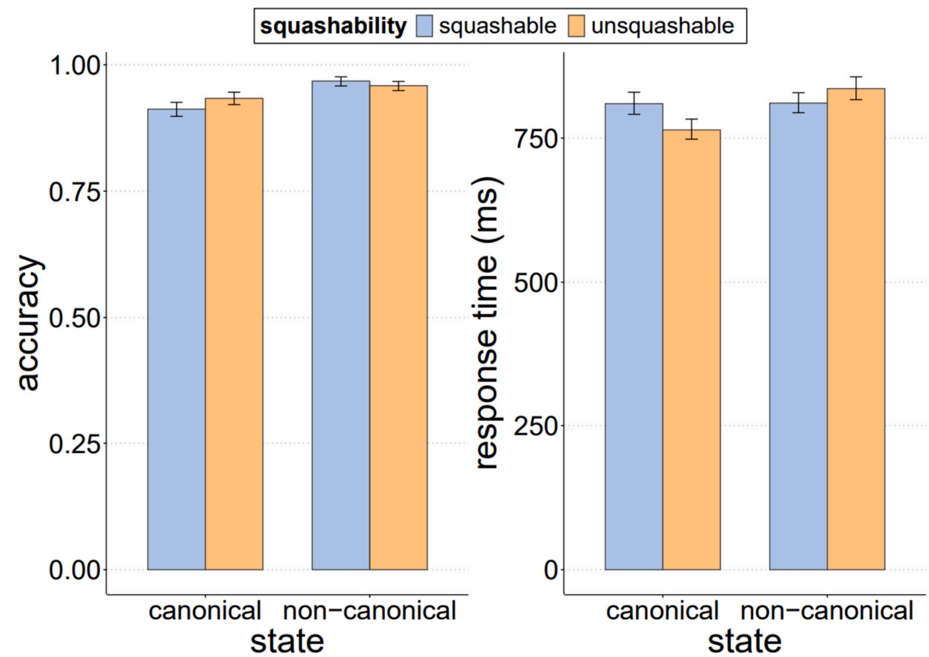

Fig. 5. Experiment 3 Results. Raw accuracy and response times as a function of object state and target object squashability in Experiment 3. Error bars represent bootstrapped 95% confidence intervals.

was higher for unsquashable items in a squashed state. Related to this issue, a recent eye-tracking study by Krass et al. (2025) found that during online comprehension of prospective sentences (e.g., *The pedestrian will open the umbrella*), there were longer and more frequent anticipatory gazes towards images depicting the consequences of the actor's goal (e.g., an open umbrella) than images of objects that merely afforded such actions (e.g., a closed umbrella). However, given that the stimuli from the present Experiment 3 represented "impossible" outcomes, the higher accuracy observed for unsquashable objects in a squashed state more likely reflects a surprisal response to an atypical stimulus.

These results, especially the lack of a match effect in response times, suggest that an object's inherent ability to be deformed was not integrated into the mental model of the event. It remains unclear if this is unique to the property of *squashability* or if target object properties are integrated in a way not measured by the sentence-picture match effect. The results nevertheless call for further investigations that extend beyond the typical "squashing" manipulation. Notably, similar to the prior experiment, a high degree of variance was captured by the random effect structure of the accuracy model that included a by-participants random slope for the interaction between squashability and state, which provided a significantly better fit than simplified models, $\chi^2(9) = 159.42$, $p < .001$. Further analysis of model variance indicated that a very high proportion of the total variability was captured by the random effects structure ($ICC = .868$), again suggesting large individual differences in the match effect. This pattern is in line with prior findings of individual differences in the ACE. Specifically, it has been posited that individual variation in response strategies, inhibitory control, and visual attention may underlie inconsistent match effects (Morey et al., 2022; Winter et al., 2022).

Experiment 4a: Sentence Focus

The first study in this series attempted to replicate Horchak and Garrido (2021), Experiment 1 in English. Despite following their procedure and using the same (translated) stimuli, we failed to replicate their finding that objects in non-canonical states were verified faster after sentences that used implicit object properties to imply state change. This was unexpected given that the original work replicated the finding over a number of variations. Experiments 2 and 3 also failed to observe a significant match effect when using a *slicing* verb to imply state change and when the target objects differed in their intrinsic *squashability*. It appears that participants in our experiments either (1) did not activate the relevant properties of the described objects or (2) did not use these properties to form altered representations of the target objects. Why might readers not be attending to this information? We next turned to examine the major factor that distinguished our experiments from the original: language.

The original study was conducted in Portuguese. Although we used the translated equivalents of the sentence stimuli provided by the authors, the syntactic structure varied across the two versions of the stimuli. Specifically, (European) Portuguese is considered a *null subject language*, whereas English requires a pronominal subject to be expressed in all finite clauses (Cognola et al., 2018; Soares et al., 2019). Consider the two sentences below:

(1) Portuguese: *Deixas cair uma bola de bowling num tomate.*
(2) English Translation: *[You] drop a bowling ball on a tomato.*

In the Portuguese version, the sentence begins with the conjugated form of the verb *deixar* ("to let [fall]"), whereas the English translation begins with the explicit subject pronoun *You*. That is, the pronoun *you* is not necessary in Portuguese (as it is implied by the conjugation), but it is grammatically obligatory in English. Importantly, presenting an explicit subject prior to the verb, as in the English sentence above, may influence what readers consider to be the focal noun phrase of the sentence (Frascarelli, 2007; Serratrice, 2008; Soares et al., 2019). Thus, the *exclusion* of the pronominal subject in the original Portuguese experiments may have unconsciously shifted focus to the acting and target objects of the sentence, rendering them more prominent in readers' mental models. This interpretation is supported by the findings of Jin et al. (2024), in which adding a focus marker to the preverbal subject (rather than the target object) completely eliminated the sentence-picture match effect for Mandarin speakers. The specific lexemic form of the verb phrase may also contribute to cross-linguistic differences. Specifically, in Portuguese, *deixas cair* uses two verbs within a periphrastic construction rather than the monolexemic form "dropped" used in English.

In our English replication of Horchak and Garrido (2021), readers may have placed preferential focus on the beginning noun phrase *You dropped* rather than the subsequent modifying phrase *a bowling ball on a tomato,* thereby reducing the saliency of the target and acting objects in their mental model and precluding a match effect. To assess the possible role of sentential focus more directly, we conducted a follow-up experiment in which we manipulated whether the target object was presented as the pre-verbal subject of the sentence or the post-verbal direct object. The experiment was preregistered (https://osf.io/6bwps). To preview the results, we observed a significant sentence-picture match effect on response times, but only when the target object was the focus of the sentence. To reconcile these findings with the null results from Experiment 1, we conducted a follow-up norming survey with a sample of native English and (European) Portuguese speakers to assess whether cross-linguistic differences in syntactic structure might account for the mixed results (see Appendix B for full details).

Method

Participants

We again conducted a simulation-based power analysis using the *simr* package (Green & MacLeod, 2016). The simulations indicated that a sample size of approximately 150 was needed to achieve 80% power. While attempting to account for future exclusions, data were collected concurrently from Prolific ($N = 200$) and an introductory psychology course at Bowdoin College ($N = 39$), resulting in a total sample of 239 participants. Ten non-native English speakers were excluded from analyses. There were no exclusions for inadequate comprehension. The final sample for the accuracy analyses was 229 participants. After excluding 43 participants for low accuracy on experimental trials, the final sample size for RT analyses was 186.

Materials and design

The design largely mirrored the first experiment with the addition of sentence focus as a factor. To account for the added factor and ensure adequate statistical power, we increased the number of target stimuli from 24 to 40. Additional target stimuli were taken from other experiments (2–7) in Horchak and Garrido (2021), and were all transformable objects (e.g., *blueberry, teapot, bottle*, etc.). There were 88 total trials, with 40 experimental trials and 48 fillers. Twenty experimental sentences were *heavy* sentences, and the other 20 were *light* sentences. Within each weight condition, 10 sentences were presented with objects in their *canonical* state and 10 in their *non-canonical* state. Critically, of the trials within each *weight* × *object state* condition, five sentences placed the target object as the focus (subject) of the sentence and five sentences placed it as the direct object at the end of the sentence. Eight randomly assigned lists were used to counterbalance the condition in which each object was presented across participants. Sample stimuli can be seen in Fig. 6.

Three additional changes were made to the format of the sentence stimuli. First, we used *falling brick* and *falling feather* (instead of *bowling ball* and *balloon*) from Experiment 6 in Horchak and Garrido (2021), as they were more commonly used items with more intuitive weights. Second, we used a sentence structure inspired by Horchak & Garrido's (2021) Experiment 6 to instantiate the sentence focus manipulation. Within this structure, we substituted a passive verb phrase that could be used with either weight or focus condition (*A… came in contact with a…*). It was necessary to change the verb from the original "drop" construction given that *A tomato was dropped on a bowling ball* results in a substantially different interpretation of the event compared to *A bowling ball was dropped on a tomato*. The primary reason we chose the "contact" construction was to enable a neutral focus manipulation that could place the target object as the subject of the sentence without directly implying a state change. Although we explored other variants, there are limited possibilities for achieving this goal within English. For example, sentence focus could also be manipulated by changing *The tomato was dropped onto the floor* to *You dropped a tomato on the floor,* but this construction does not manipulate implicit properties of an acting object. Likewise, more forceful verbs, such as *smashed*, are not appropriate because they imply state change regardless of the specific acting object.

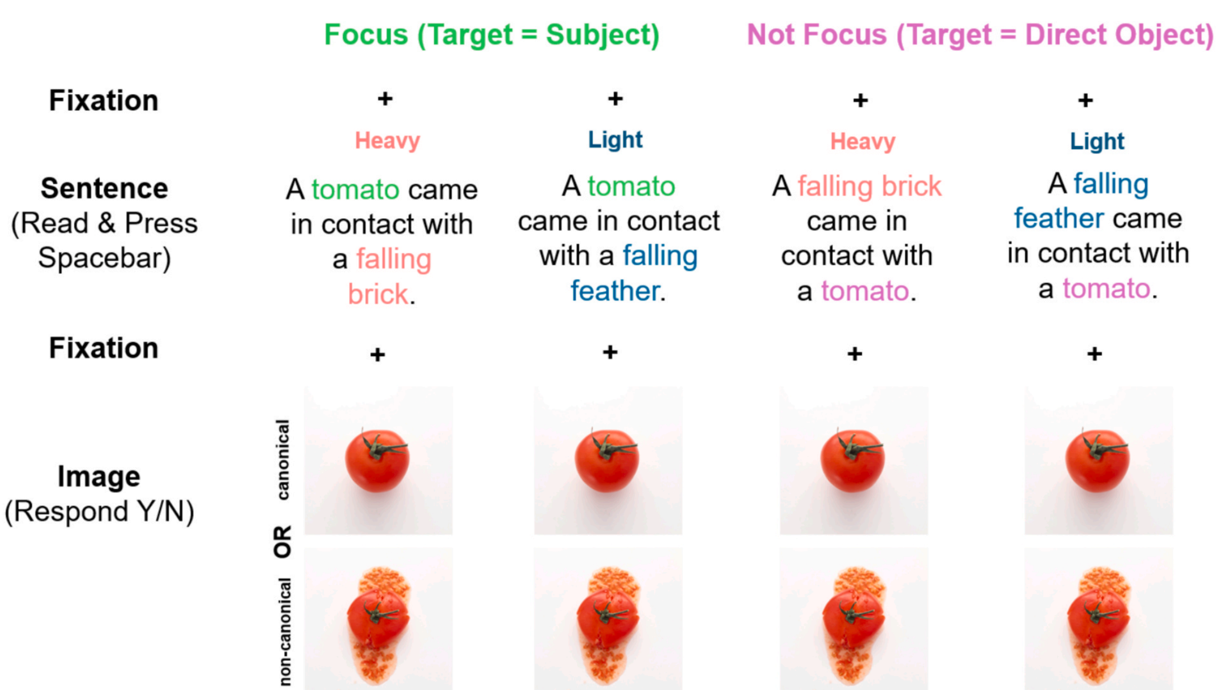

Fig. 6. Overview of Experiment 4a Procedure. Target stimuli conditions, examples, and participant interactions in Experiment 4a. Sentences manipulated the weight of the falling object (heavy vs. light) and the focus of the sentence (target either subject or direct object). The object in the image stimulus was either in its canonical (i.e., intact) or non-canonical (i.e., smashed) state.

Thus, we settled on the *contact* construction because it adequately accommodated the manipulation of both sentence focus and falling object weight (but see Experiment 4b). Filler trials were adapted to follow a similar format.

Finally, to minimize encoding of target objects outside of the experimental trial context, we replaced any target items used as lures in filler trials with unrelated non-target objects. In an effort to more closely approximate the original study, we once again included comprehension questions in the same format as Experiment 1.

Procedure

The procedure was identical to Experiment 1 with the exception that participants used the 'Y' key for 'yes' responses and 'N' for 'no' responses during the task. Data were processed as in the previous experiments.

Results

Accuracy analysis

Mean accuracy was 88.38%. We fit a logit mixed-effects model using weight, object state, sentence focus, and their interactions as fixed effects; random intercepts for participants and items; and by-participant random slopes for the main effects of weight, object state, and sentence focus (the model did not converge when including the interaction term). There was a main effect of weight ($b = 1.23$, $z = 4.52$, $p < .001$), such that accuracy was higher following light compared to heavy sentences. There was also a main effect of object state ($b = 2.23$, $z = 7.35$, $p < .001$), such that accuracy was higher for pictures of non-canonical (smashed) objects. There was no main effect of sentence focus ($b = 0.08$, $z = 0.39$, $p = .695$). As shown in Fig. 7 (left panel), there was a significant two-way interaction between weight and object state ($b = -3.61$, $z = 11.15$, $p < .001$), such that responses were more accurate when canonical objects were preceded by light sentences and when non-canonical objects were preceded by heavy sentences. The three-way interaction was not significant ($b = -0.11$, $z = -0.29$, $p = .771$). Thus, in contrast with both Horchak and Garrido (2021) and Experiments 1–3, we observed a sentence-picture match effect on accuracy measures regardless of sentence focus.

Response time analysis

We fit a linear mixed-effects model to the transformed RT data, with the same fixed and random effects structure described for the accuracy model. There was a significant main effect of weight ($b = -0.02$, $t = -3.39$, $p < .001$), such that response times were faster after light sentences. There was also a main effect of sentence focus ($b = -0.02$, $t = -2.50$, $p = .012$), in which response times were faster following sentences that placed the target object as the focal subject. There was a significant two-way interaction between weight and object state ($b = .04$, $t = 4.17$, $p < .001$). However, as shown in Fig. 7 (right panel), this was qualified by a significant three-way interaction with sentence focus ($b = .03$, $t = 2.75$, $p = .006$). Follow-up contrasts indicated that non-canonical objects were verified faster following heavy sentences ($b = 0.02$, $t = 2.50$, $p = .013$) as in Horchak and Garrido (2021), but *only* when the target object was the focus of the sentence. In contrast to Horchak and Garrido (2021), canonical objects were also verified faster following light compared to heavy sentences ($b = 0.02$, $t = 3.39$, $p < .001$), but again only when the target object was the focus of the sentence.

Follow-up norming survey

These results suggest that the sentence-picture match effect is only found when the target object is appropriately emphasized in the mental model. Given that (European) Portuguese (the language used by Horchak & Garrido, 2021) typically utilizes a null subject construction, we reasoned that the target object may be more salient in Portuguese than in the English equivalent. The results of Experiment 4a appear to support this interpretation. To assess our interpretation more directly, we conducted a brief, pre-registered follow-up survey with both native English and Portuguese speakers. The survey assessed how readers interpret the events described by the different verb phrases used across experiments and the relative salience of the described objects (see Table 2). Here we report the key findings. A full description of the methods and results can be found in Appendix B.

Method. Across populations, we sought to investigate three key questions. First, we investigated how participants interpreted sentences that used the format from Experiment 1 (*You drop a bowling ball on a tomato*) versus Experiment 4a (*A falling brick came in contact with a tomato*). Second, we assessed whether the tense of the verb (*drop/dropped* vs. *comes/came in contact*) further modulates the interpretation of the event. Lastly, within the "Contact" condition, we assessed the impact of the focus manipulation by mirroring the structures used in Experiment 4a. For the Portuguese sample, we added another manipulation. Specifically, within the "Drop" conditions, we also manipulated whether the sentence used a null or explicit subject construction (e.g., *Deixas cair uma bola de bowling num tomate* vs. *Tú deixas cair uma bola de bowling num tomate*).

Two separate surveys were constructed for English and Portuguese speakers. Participants answered four questions about each sentence construction (see Table 2), but here we focus on the results from the final question where participants were asked to rank the nouns in the

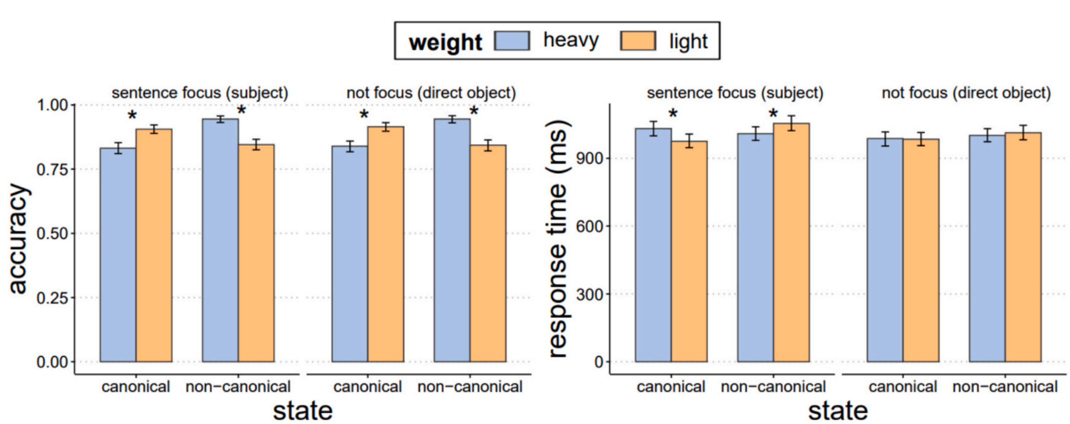

Fig. 7. Experiment 4a Results. Mean accuracy and response times (RTs) in Experiment 4a. Accuracy was higher for non-canonical objects following heavy sentences and for canonical objects following light sentences, regardless of sentence focus. RTs to non-canonical objects were faster following heavy sentences, but only when the target object was the focus (subject) of the sentence. Contrary to Horchak and Garrido (2021), RTs to canonical objects were faster after light sentences, but again, only when targets were the focus of the sentence. Error bars represent bootstrapped 95% confidence intervals and asterisks indicate a significant effect ($p < .05$).

Table 2
Sentence Stimuli.

Experiment	Focus Condition	Example Sentence	Norming
1		*You dropped a bowling ball on a tomato.*	Experiment 4a and 4b
4a	Not Focus	*A falling brick came in contact with a tomato.*	Experiment 4a and 4b
4a	Focus	*A tomato came in contact with a falling brick.*	Experiment 4a and 4b
4b	Not Focus	*You dropped a bowling ball onto a tomato.*	Experiment 4b
4b	Focus	*Onto a tomato you dropped a bowling ball.*	Experiment 4b

Questions
1) Before the described event occurred, the [object] was…
2) After the described event occurred, the [object] was…
3) Does the sentence imply that the [object] was smashed?
4) When you imagined the event unfolding, how prominent was each of the following items in your mental image of the event?

The full details of the norming studies from Experiments 4a and 4b can be found in Appendix B and Appendix C, respectively.

sentence in order of salience (see Appendix B for additional results). Specifically, we asked the following:

> *"When you imagined the event unfolding, how prominent was each of the following items in your mental image of the event? Click and drag to reorder the items by prominence. Ranking an item as #1 indicates that it played the most prominent role in your mental image (*i.e.*, the item was the focus of your "mental spotlight")."*

For "Drop" sentences, participants ranked the *falling object*, the *target object*, and *you*. For "Contact" sentences, they only ranked the *falling* and *target objects*.

Verb and Focus-Related Effects. Among English speakers, for "Drop" sentences (e.g., "You dropped…"), the falling object was most likely to be ranked as the focal (most prominent) item, $b = .82$, $z = 2.81$, $p = .004$. For "Contact" sentences ("A… came in contact with a…"), the falling object was most likely to be ranked as the focal item in the Not Focus condition (e.g., "A falling brick came in contact with a [tomato]"), whereas the target object was most likely to be ranked as the focal item in the Focus condition (e.g., "A [tomato] came in contact with a falling brick"), $b = 3.61$, $z = 6.97$, $p < .001$.

For Portuguese speakers, the results for both "Drop" sentences ($b = 1.05$, $z = 3.41$, $p < .001$) and "Contact" sentences ($b = 3.19$, $z = 6.34$, $p < .001$) mirrored those for English speakers. When examining the null vs. explicit subject manipulation, contrary to our predictions, the target object was *not* ranked differently across null and explicit subject conditions, $b = .01$, $z = .03$, $p = .979$. However, "you" had a significantly higher probability of being ranked as the focal item in the explicit subject condition, $b = 1.54$, $z = 2.96$, $p = .003$.

Cross-Language Comparisons. In European Portuguese, null subject constructions are far more common than explicit subject constructions. Therefore, the comparison between these constructions in Portuguese speakers may have been clouded by the sheer atypicality of the explicit construction. To better understand the differences in object saliency in null and explicit subject constructions, we compared how Portuguese speakers ranked items in the null subject "Drop" condition to how English speakers ranked items in "Drop" sentences. That is, we directly contrasted the original sentences used in Horchak and Garrido's (2021) Experiment 1 with the provided English translations we utilized in the present Experiment 1. Given that there were no significant differences between present and past tenses in prior analyses, we collapsed across tenses for both groups. Crucially, as shown in Fig. 8, we found that the target object was significantly more likely to be ranked as the focal item by Portuguese speakers than English speakers, $b = .50$, $z = 2.31$, $p = .021$. The same pattern was observed for the falling object, $b = .72$, $z = 3.11$, $p = .001$. In contrast, "you" was more likely to be ranked as the focal item by English speakers than Portuguese speakers, $b = 1.12$, $z = 4.22$, $p < .001$.

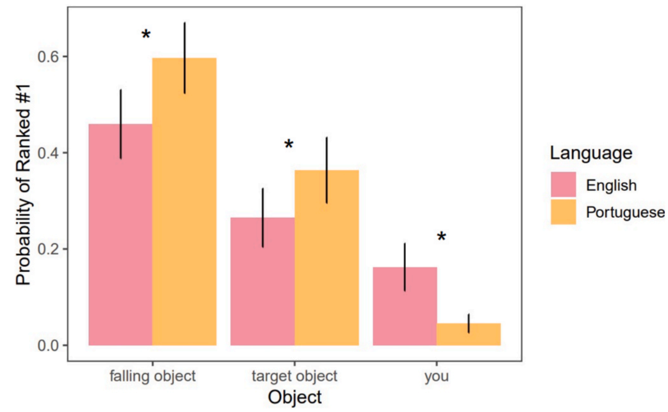

Fig. 8. Ranking Results. Probability of ranking the *falling object*, *target object*, and *you* as the most prominent item in the sentence. Portuguese speakers were more likely than English speakers to rank the *falling object* and the *target object* as the most prominent item, whereas English speakers were more likely than Portuguese speakers to rank "you" as the most prominent item. Error bars represent bootstrapped 95% confidence intervals, and asterisks indicate a significant effect ($p < .05$).

Discussion

In the present experiment, we sought to clarify the null results of Experiment 1 and our subsequent failure to find significant match effects in Experiments 2 and 3 by investigating how syntactic factors may modulate the sentence-picture match effect. Specifically, we investigated whether this effect depends on the target object being appropriately salient in the mental model. We hypothesized that the explicit subject construction used in English may have drawn attention to the pronominal subject at the expense of the described objects. We therefore manipulated the sentence stimuli to place the target object as either the focal subject or the direct object in the sentence. We found that objects in modified states were verified faster after sentences implying state change, but only when the target object was the focus of the sentence. A follow-up survey of native English and Portuguese speakers found that the surface structure of the Portuguese stimuli led to more salient representations of the target object, supporting our hypothesis that the sentence-picture match effect depends on a syntactic structure that appropriately accentuates the target object (Jin et al., 2024).

In contrast to Horchak and Garrido (2021), we also found that *canonical* objects were verified faster following sentences implying minimal state change (i.e., light sentences), but again only when the target object was the focal subject. This finding challenges the interpretation that both initial and modified object states remain accessible during comprehension (Altmann & Ekves, 2019; Horchak & Garrido, 2021). Our results suggest instead that it is primarily the final object state that is represented in the mental model of the event.

Lastly, we observed a sentence-picture match effect for accuracy, regardless of sentential focus. No such effect was observed in Experiment 1 or in Horchak and Garrido's (2021) Experiment 6, which used a similar structure to our direct object ("not focus") condition. This effect may have been driven by the different expectancies generated by the *contact* construction. Another possibility is that there was simply more room for accuracy effects to manifest, as the average accuracy was somewhat lower in Experiment 4a (88.38%) than Experiment 1 (95.24%), likely due to the increased number of experimental trials. More work is needed to disentangle these possibilities.

In this cross-experiment comparison, it is important to consider the potential impact of changing the verb phrase. Recall that Experiment 1 used the directional construction *You dropped a… on a…*, while Experiment 4a used the neutral construction *A falling… came in contact with a…* to allow the simultaneous manipulation of falling object weight and sentence focus. Although the sentences using these verb forms express a similar scenario of two objects interacting with each other, the verbs have nuanced interpretations. Specifically, "drop" implies a direct causal effect, in which the acting object is necessarily above the target object before being dropped, and is also intuitively associated with a greater degree of state change. In contrast, "came in contact with" implies less force and is not inherently directional, such that the agency of the acting object is drawn from the adjective "falling" rather than the verb itself. Consistent with this interpretation, the survey results indicate that both English and Portuguese readers felt that the *contact* construction implied a lesser degree of state change relative to the *drop* construction. However, admittedly, this divergence makes it difficult to isolate the effects of the sentence focus manipulation.

Nevertheless, the results of Experiment 4a cannot be fully attributed to this verb change because the same verb construction was used across *both* focus conditions, yet a sentence-picture match effect on response times was observed *only* when the target object was the focus. It is also possible that the change in verb interacted with the focus manipulation. Recall that we observed a sentence-picture match effect across both focus conditions on accuracy measures, but only observed an effect on response times within the focus condition. The weaker implications of "came in contact with" may only yield a match effect on response times when focus is explicitly placed on the target object.

Another important consideration is the familiarity of the target objects across experiments. To appropriately power the added focus manipulation in Experiment 4a, we increased the number of target items by using additional stimuli from Experiments 2–7 in Horchak and Garrido (2021). It is possible that the stimuli from Experiment 1 were more familiar to Portuguese speakers, making it easier to represent these objects in mental models during comprehension. The added stimuli in Experiment 4a may have introduced more familiar items to the English sample, thereby amplifying the match effect. However, given that this effect was only found in the focus condition for response times and the assignment of items to conditions was fully counterbalanced, this interpretation is unlikely. Further, we examined the average frequency of the target object nouns in Experiment 1 in both Portuguese (SUBTLEX-PT; Soares et al., 2015) and English (SUBTLEX-US; Taylor et al., 2020). Items that were not included in the databases (e.g., iPhone) were excluded. For items with data available across both languages, the noun frequencies were similar in Portuguese ($M = 4.04$, $SD = .53$) and English ($M = 4.17$, $SD = .85$). This analysis suggests that familiarity cannot account for the present findings. Related to this issue, as in Experiment 1, we conducted follow-up analyses that included ratings of how well a given smashed object matched what participants imagined when reading heavy sentences as a covariate (see Appendix C). The results mirrored those reported above. Thus, although more work is needed to fully understand the specific conditions under which the match effect occurs, the current pattern of results suggests that sentence focus is a likely moderator.

Experiment 4b: sentence focus replication

In Experiment 4a, we found that objects in modified states were verified faster after sentences implying state change, but only when the target object was the focus of the sentence. This key interaction, paired with the results from our subsequent survey of English and Portuguese speakers, suggests that the sentence-picture match effect depends on a syntactic structure that appropriately accentuates the target object (Jin et al., 2024). However, the use of a markedly different syntactic structure in Experiment 4a complicates direct comparison to Experiment 1. The stimuli used in Experiment 4a (e.g., "A falling brick came in contact with a tomato") only contained two nouns (the falling object and the target object), whereas the stimuli from Experiment 1 (e.g., "You dropped a bowling ball on a tomato") also included the pronominal subject "you," potentially clouding the interpretation of sentential focus. Moreover, the survey results of Experiment 4a showed that the sentence stimuli used in Experiment 4a were judged to imply less state change than the sentence stimuli used in Experiment 1, suggesting that sentence focus may have been confounded with causal strength. We therefore sought to replicate Experiment 4a with modified sentence stimuli that more closely approximated those used in Experiment 1.

To again preview the results, we failed to replicate the moderating effect of sentence focus and did not observe any significant sentence-picture match effects on response times, although we did find a match effect on accuracy once again. We reasoned that lingering differences in noun saliency and perceived causality might still be driving these results, and thus conducted a second follow-up norming study to gain insight into how participants parsed sentences with different structures (see Results section and Appendix C).

Method

Participants

Sample size estimates were based on the power analysis conducted for Experiment 4a. 174 participants were recruited from Prolific. Five non-native English speakers were excluded from analyses. Two participants were excluded for inadequate comprehension. The final sample for the accuracy analyses was 167 participants. After excluding 10 participants for low accuracy on experimental trials, the final sample size for RT analyses was 157.

Materials, design, and procedure

The picture materials, design, and procedure directly mirrored those of Experiment 4a, with the only modification being the structure of the sentence stimuli. The stimuli were modified to be as similar to the Experiment 1 stimuli as possible while still allowing for the focus manipulation. The conditions utilized the following format:

Focus Condition: *Onto a tomato you dropped a bowling ball.*
Not Focus Condition: *You dropped a bowling ball onto a tomato.*

Results

Accuracy analysis

Mean accuracy was 94.23%. The structure of the logistic mixed-effects model was the same as in Experiment 4a. As in Experiment 4a, there was a main effect of weight ($b = 0.77$, $z = 2.50$, $p = .013$), where accuracy was higher for light compared to heavy sentences. There was again a main effect of object state ($b = 1.00$, $z = 2.98$, $p = .003$), with accuracy higher for pictures of non-canonical (smashed) objects. As shown in Fig. 9 (left panel), there was again a significant interaction between weight and object state ($b = -1.72$, $z = -4.35$, $p < .001$), where canonical objects were verified more accurately after light sentences and non-canonical objects were verified more accurately after heavy sentences. There was no main effect of sentence focus ($b = -0.10$, $z = -0.42$, $p = .678$), and sentence focus did not significantly interact with weight ($b = -0.09$, $z = -0.27$, $p = .786$) or object state ($b = 0.36$, $z = 0.91$, $p = .363$). Thus the results of the accuracy analysis were largely consistent with Experiment 4a.

Response time analysis

The structure of the linear mixed-effects model was the same as in Experiment 4a. As shown in Fig. 9 (right panel), there was no main effect of weight ($b < .001$, $t = -0.13$, $p = .897$). However, there was a main effect of object state ($b = 0.02$, $t = 2.47$, $p = .013$), with objects in canonical states verified faster than objects in non-canonical states. Surprisingly, contrary to Experiment 4a, there was no main effect of

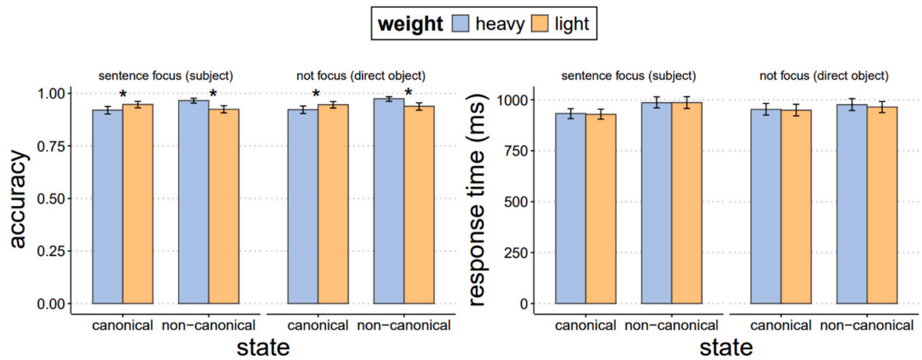

Fig. 9. Experiment 4b Results. Mean accuracy and response times (RTs) in Experiment 4b. Accuracy was higher for non-canonical objects following heavy sentences and for canonical objects following light sentences, with no effect of sentence focus. In contrast to Experiment 4a, there were no significant effects in the RT analyses. Error bars represent bootstrapped 95% confidence intervals, and asterisks indicate a significant effect ($p < .05$).

sentence focus ($b < .001$, $t = 0.04$, $p = .970$), and sentence focus did not significantly interact with weight ($b < .001$, $t = 0.03$, $p = .974$) or object state ($b = -0.01$, $t = -0.43$, $p = 0.671$). Thus, the results of the present experiment are in line with our initial failure to replicate Horchak and Garrido (2021) in Experiment 1, but deviate from the findings of Experiment 4a.

Follow-up norming survey

To gain more insight into the factors that contributed to the different sentence-picture match effects observed across Experiments 4a and 4b, we collected normative information about the sentence and picture stimuli used across experiments. For the sentence stimuli, we collected participant feedback regarding what they mentally simulated when reading the sentences from each experiment. Here we report the key findings from the sentence norming portion relevant to Experiment 4b. The full details and report of this survey can be found in Appendix C.

Design, Materials, and Procedure. Participants were presented sentences with different syntactic structures and asked questions about their mental simulation of the described events. Participants completed one trial for each of the sentence structures used across experiments; see Table 2.

As in the Experiment 4a norming study, the target object (tomato, blueberry, or muffin) was randomly selected for each condition. On each trial, participants were presented with a sentence. After five seconds, four questions with sliding scales (0–100 range) appeared (see Appendix C). These questions were similar to those from the Experiment 4a norming study, but here we focus on the two most informative questions:

1) *Does the sentence imply the [target object] was smashed? (0 = Not at All, 100 = Definitely Yes)*
2) *When you imagined the event unfolding, how prominent was each of the following items in your mental image of the event? Rating an item highly indicates that it played a prominent role in your mental image (*i.e., *the item was the focus of your "mental spotlight") (0 = Not Prominent At All, 100 = Extremely Prominent)*

Does the sentence imply the [target object] was smashed? Sentences from Experiment 1 were judged to more strongly imply state change than the Focus ($b = -28.73$, $t = -5.82$, $p < .001$) and Not Focus ($b = -22.63$, $t = -4.59$, $p < .001$) conditions from Experiment 4a, consistent with our findings from the previous norming study reported in Appendix B. Critically, however, sentences from Experiment 1 implied a similar degree of stage change as the Focus ($b = -3.89$, $t = -0.79$, $p = 0.431$) and Not Focus ($b = -1.31$, $t = -0.27$, $p = 0.790$) conditions from Experiment 4b. Taken together, these findings suggest that, as expected, the syntactic structures used in Experiment 4a implied less extreme state change and less direct causality than the structures used in Experiments 1 and 4b, which were similar to each other on these dimensions.

When you imagined the event unfolding, how prominent was each of the following items in your mental image of the event? Rating an item highly indicates that it played a prominent role in your mental image (i.e., the item was the focus of your "mental spotlight"). Note that in the present survey, participants provided a quantitative saliency rating for each noun (i.e., falling object, target object, and "you" when relevant) rather than rank-ordering them (as in the previous norming study reported in Appendix B). We first examined the saliency of the falling object. Compared to the structure used in Experiment 1, the falling object was rated to be similarly prominent in the Experiment 4a Focus condition ($b = 3.91$, $t = 0.78$, $p = .435$) and marginally more prominent in the Not Focus condition ($b = 9.08$, $t = 1.82$, $p = .070$); see Fig. 10. Similarly, compared to Experiment 1, the falling object was rated to be similarly prominent in the Experiment 4b Focus condition ($b = -6.11$, $t = -1.22$, $p = .222$) and marginally more prominent in the Not Focus condition ($b = -9.25$, $t = -1.85$, $p = .065$).

Critically, we reasoned that the saliency of the simulated target object was likely the most influential factor in observing a significant sentence-picture match effect. Compared to the structure used in Experiment 1, the target object was marginally more salient in the Experiment 4a Focus condition ($b = 9.47$, $t = 1.85$, $p = .065$) and significantly more salient in the Not Focus condition ($b = 11.11$, $t = 2.17$, $p = .030$). There were no other significant differences across other sentence structures. We also compared the saliency of target objects in Experiments 4a and 4b. While target objects in the Focus condition in Experiment 4a were similarly salient to those in the Focus condition in Experiment 4b ($b = -5.67$, $t = -1.11$, $p = .268$), target objects in the Not Focus condition in Experiment 4a were significantly more salient than those in the Not Focus condition of Experiment 4b ($b = -11.75$, $t = -2.30$, $p = .022$).

Discussion

In the present experiment, we attempted to replicate the results of Experiment 4a with materials that more closely approximated the original stimuli used in Experiment 1. While we observed the same sentence-picture match effect on accuracy found in Experiment 4a, we unexpectedly failed to observe a significant sentence-picture match effect on response latencies, regardless of sentential focus. This suggests that other factors beyond the manipulated noun order contributed to the match effect observed in Experiment 4a. Follow-up analyses including normative ratings as a covariate did not alter these patterns (see Appendix C).

The materials in Experiment 4a and the present experiment differed in two key ways. First, as supported by the results of both surveys, the verb phrase "came in contact with" from Experiment 4a implies less

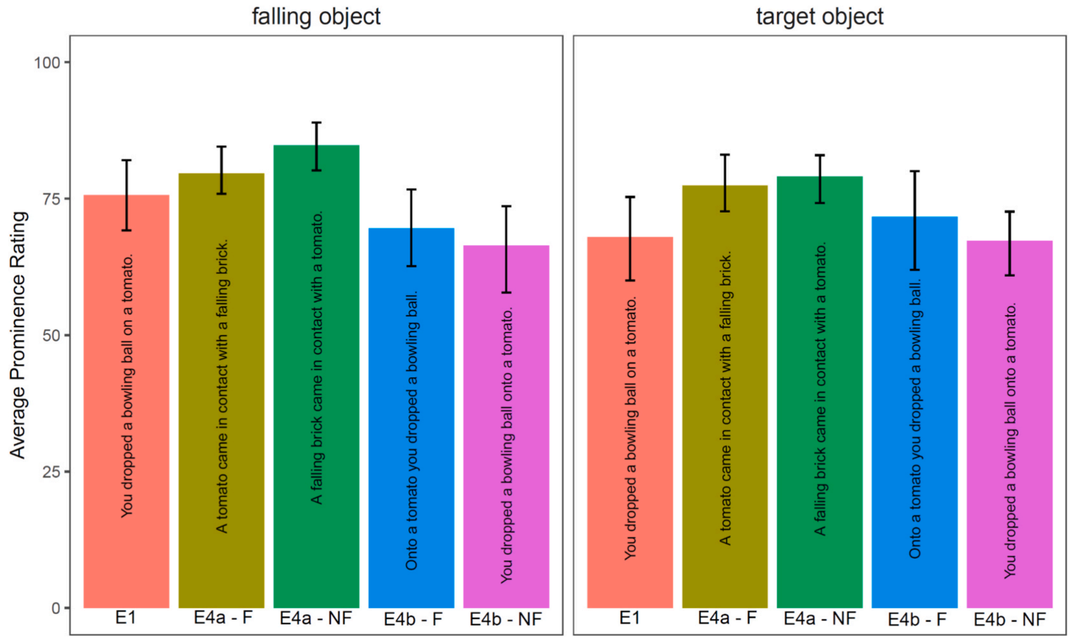

Fig. 10. Prominence Ratings. Prominence ratings for the falling and target objects across seven key sentence structures. Example sentences are overlaid on each bar. E1 = Experiment 1, E4a-F = Experiment 4a Focus condition, E4a-NF = Experiment 4a Not Focus condition, E4b-F = Experiment 4b Focus condition, E4b-NF = Experiment 4b Not Focus condition. Error bars represent bootstrapped 95% confidence intervals.

deformation and less direct causality than the "dropped" structure used in Experiment 1 and the present experiment. Second, while there was some degree of variation, the target object was often more salient than the falling object in both conditions of Experiment 4a. It is important to note that in addition to the more ambiguous causal structure, the phrasing in Experiment 4a may have been syntactically less complex compared to other experiments, thereby simplifying simulation demands.

We suggest that the present results are due to these two intertwined factors. It might seem that sentences with a clear causal structure would elicit more robust simulations during comprehension. However, if this were the case, one would not expect the match effect we observed for the ambiguously causal stimuli of Experiment 4a. Therefore, it is possible that sentences with weaker causality actually elicit *stronger* mental simulation as comprehenders attempt to disambiguate the outcome of the described event. In addition, the sentences in Experiment 1 and the present experiment included three nouns: the falling object (balloon/bowling ball), the target object, and the pronominal subject "you." By contrast, the sentences in Experiment 4a only included the falling and target objects. In line with past work (e.g., Gordon et al., 2004; see also Frank & Vigliocco, 2011), increasing the number of nouns in the comprehended sentence may have diluted the saliency of *all* described entities, thereby weakening the representation of state change. We elaborate on these issues in the General Discussion.

General discussion

Accumulating evidence suggests that mental models are used to simulate described events during language comprehension (Glenberg et al., 1999; Glenberg et al., 1999; Horchak and Garrido, 2024; Zwaan & Pecher, 2012). This process entails activating both surface-level and implicit properties of objects and tracking how interactions between objects influence the state of a given object (Altmann & Ekves, 2019). In recent work, Horchak and Garrido (2021) found that after reading sentences in which a heavy (versus light) object was dropped onto a malleable target object, participants were faster to verify that a smashed, non-canonical depiction of the object had been mentioned in the preceding sentence. In the present work, we sought to replicate this initial finding and offer several clarifying extensions. Below, we summarize the key patterns across all experiments.

Experiment 1 was an English replication of the first study by Horchak and Garrido (2021), but we failed to observe the same sentence-picture match effect. Experiment 2 pivoted to examine a different type of state change, *slicing,* but we again failed to observe any benefit for verifying objects whose state matched the event described in the previous sentence. Experiment 3 assessed the representation of implicit properties of the *target* object (rather than the acting object) by manipulating whether the target object was inherently *squashable*. Again, we observed no match effect for squashable objects in a non-canonical state after sentences describing a state change event. Finally, in Experiment 4 we investigated how the surface structure of the sentence may moderate the representation of described objects, with the goal of shedding light on the initial replication failure and the absence of the match effect in the extensions. We posited that the null subject form of Portuguese may have directed relatively more attention to the acting and target objects in the original study, thus accentuating their representation in the mental model compared to English. We examined this possibility by manipulating whether the target object was the focus (subject) of the sentence or appeared in the sentence-final position (direct object). Responses to objects in non-canonical states were indeed facilitated when sentences with underspecified causal structures (Experiment 4a, "came in contact with") were used to imply state change, but only when the target object was the focus of the sentence (e.g., *A tomato came in contact with a falling brick*). In contrast, sentences with more direct causal structures, similar to those used in Experiment 1, did not elicit a sentence-picture match effect, regardless of sentence focus (Experiment 4b; e.g., *Onto a tomato you dropped a bowling ball*). To clarify these findings, we conducted follow-up norming surveys with both English and Portuguese speakers to better understand how these sentence structures are interpreted. The results indicated that Portuguese speakers viewed the target and acting objects as more salient than English speakers (Appendix B), and that, in a separate sample of English speakers (Appendix C), target objects were typically less salient in structures with an explicit pronominal subject (i.e., "you"). Taken together, the body of evidence suggests that the simulation of object state change during comprehension is not automatic or ubiquitous, but rather is sensitive to the joint influence of causal structure and surface-level syntactic properties.

Representation of object states

When using sentences with clear causal structures (e.g., *You dropped a bowling ball on a tomato*), we failed to find a benefit for verifying objects in a non-canonical state after reading sentences implying significant state change. These null results were not limited to the interaction of falling object weight and object state, as originally examined by Horchak and Garrido (2021), but were also found when examining *slicing* state change and the *squashability* of target objects. These findings cast doubt on the claim that objects are automatically transformed within mental models during comprehension, at least in a way that is readily measurable via the sentence-picture match effect. However, we did observe faster verification latencies for objects in a non-canonical state after heavy sentences and for objects in a canonical state after light sentences in Experiment 4a, but only when the target object was the subject of a sentence with an underspecified causal structure (e.g., *A tomato came in contact with a falling brick*).

Also, in both Experiments 4a and 4b, we observed a match effect on accuracy for objects in both canonical and non-canonical states, regardless of sentence focus. Recall that Horchak and Garrido (2021) argued that both canonical and modified object states remain active, and may even compete with one another, during event comprehension, in line with the notion of intersecting object histories (Altmann & Ekves, 2019; see also Hindy et al., 2012; Solomon et al., 2015). This position was supported by the finding that the match effect in their study was only found for smashed objects following sentences implying state change, and was not found for objects in canonical states following sentences implying minimal to no state change.

More recent work by Horchak and Garrido (2024) echoes this claim. In one experiment, they replicated the match effect for objects in non-canonical states following sentences using direct verb phrases (rather than implicit knowledge of object properties) to imply state change (e.g., *Jane stepped on a banana*) and also found a match effect for canonical objects following neutral sentences (e.g., *Jane chose a banana*). Although canonical objects were verified faster overall after neutral sentences compared to sentences implying state change, there was no significant difference in verification latencies for canonical and non-canonical objects after heavy sentences when segregating the analysis by sentence type. The authors argued that the latter pattern of results indicates that both object states are similarly accessible. However, it is likely that the activation of these competing states varies based on sentential context. In a sentence-picture verification paradigm, Prystauka et al. (2024) used adverbs to modify the degree of implied state change during comprehension. They found that intensifier adverbs (e.g., *completely*) hindered the accessibility of the target object's canonical state, providing strong evidence that morphosyntactic devices can modulate the activation of object states during comprehension.

The interpretation of the match effect for objects in canonical states remains unclear and warrants further investigation, especially in cross-linguistic work. We suggest that although intact and modified states may both be active following the comprehension of a state change event, the contextually appropriate state may be much more salient (Prystauka et al., 2024). Critically, the inconsistent results of our experiments, both in comparison to Horchak and Garrido (2021) and within the present series, suggest that the sentence-picture match effect may not be a particularly robust phenomenon (also see Winter, 2022).

Forms of state change

Experiment 1 failed to find a match effect for images of squashed objects after sentences implying squashing. To better understand the conditions in which readers form salient, contextually appropriate representations of objects during comprehension, Experiment 2 moved away from the state change of *squashing* to examine how readers represent a more common form of state change, *slicing*. Rather than manipulating the weight of the falling object, sentences in Experiment 2 either implied no state change (*Someone places an apple in front of you*) or described a slicing event (*Someone slices an apple in front of you*), and pictures depicted either intact or sliced objects. Although the results trended in the expected direction, we did not observe a significant match effect for sliced objects after sentences describing slicing, nor for objects in a canonical state after neutral sentences. This was surprising given the strong implications of slicing and the fact that the study was well powered.

It is possible that attention was not adequately directed to the target object based on the sentence structure (e.g., "Someone places/slices an apple in front of you"), thereby minimizing its saliency in the mental model. In line with syntactic influences, it is also possible that a past tense format, which implies that the state change event has already occurred, would elicit stronger effects (Kang et al., 2020a; 2020b). Regardless, the present results suggest that not all forms of state change are simulated to the same degree, or at least in the same manner. They also highlight that event representation entails diverse forms of object manipulation, calling for investigations of state change beyond squashing and slicing.

Object properties

In Experiment 3, we examined implicit properties of the to-be-verified target object instead of the falling object. Our paradigm extended previous work examining intrinsic squashability using eye-tracking (e.g., Kang et al., 2020b) as well as work investigating target object color, where the color of the object (e.g., a steak) either matches (e.g., a brown, cooked steak) or mismatches (e.g., a raw, red steak) the color implied by the preceding sentence (e.g., *John looked at the steak on his plate*; Connell et al., 2007; also see Supplementary Materials). Examining target object properties is important because such properties directly determine the initial object state as well as the actual feasibility of the described modification. We specifically manipulated whether the target object was inherently *squashable,* confirmed via a norming study. For example, a *tomato* would almost certainly be smashed under the weight of a falling bowling ball, whereas the state of a *book* is unlikely to be impacted by such an event. Even though the study was adequately powered, we again failed to find a match effect for squashable objects in a deformed state after reading sentences implying state change, nor any significant differences between squashable and unsquashable objects.

There are several possible explanations for these null findings. For one, it is again possible that the structure of the sentence (e.g., "You dropped a bowling ball on a tomato / penny") directed relatively little focus to the target object. It is possible that the impact of implicit object properties may be more impactful when they are characteristic features of the object. For example, heaviness may be considered characteristic of *bowling balls* but not *printers* even though both objects are typically heavy. Thus, although *squashability* may be characteristic of a subset of the target objects used in Experiment 3 (e.g., an egg), it is unlikely to have been previously encoded as a salient property of all of the objects (e.g., sushi) in the study. In turn, *unsquashability* is unlikely to have been considered characteristic of the unsquashable objects (e.g., a book), which may have limited the constraints placed on these objects during comprehension. These observations suggest that the representation of implicit object properties, especially those of target objects, depends on how central the property is to the concept as a whole. Future work might systematically vary how characteristic a given property is of various target objects and examine whether this factor moderates sentence-picture match effects.

Syntactic constraints

The key influence of syntactic structure was demonstrated by the findings of Experiments 4a and 4b, which suggest that surface-level properties of the sentence structure can modulate how entities are represented in mental models of events. Whereas the sentence structure used in the original Portuguese work did not include an explicit pre-verbal subject, the English translations required one (*you*). As a result, Horchak and Garrido's (2021) Portuguese readers may have focused more on the described objects and incorporated the implicit properties of these objects into their mental model of the described event, whereas English readers in our first experiment may have focused more on simulating the first noun phrase of the sentence (i.e., themselves performing the act of dropping). In Experiment 4a, however, the sentence structure omitted the pre-verbal pronoun, thus directing readers' attention more to the state-change event itself. When the first noun phrase of the sentence highlighted the object that underwent the state change, participants appeared to shift their mental model to simulate the impact of the acting object, as confirmed by the significant match effect and ratings of object salience.

Yet, as discussed previously, it is fair to question whether the results of Experiment 4a are fully attributable to the sentence focus manipulation, as well as whether this experiment is comparable to the failed replication in Experiment 1. The key manipulation in Experiment 4a was whether the sentence structure placed the target object as the focal subject or the direct object by changing the central verb phrase from *You [dropped] a bowling ball on a tomato* to *A falling brick [came in contact with] a tomato*. However, we recognize that this change to the verb phrase, while necessary for the focus manipulation, complicates direct comparison to the original Horchak and Garrido (2021) study and Experiment 1, both of which used the "drop" verb format. Moreover, norming the stimuli indicated that the "came in contact" verb phrase used in Experiment 4a implied less direct causality than the "dropped" construction used in Experiment 1, which may have influenced mental simulation processes separately from sentence focus.

In Experiment 4b, we thus attempted to replicate Experiment 4a with Focus (e.g., *Onto a tomato you dropped a bowling ball*) and Not Focus (e.g., *You dropped a bowling ball onto a tomato*) conditions that more closely approximated the sentence structures used in Experiment 1. Using these constructions, we failed to observe a sentence-picture match effect on response times, regardless of sentence focus. Given that the sentence-picture match effect in Experiment 4a was only found in the Focus condition, these results suggest that the results of Experiment 4a were due to a combination of sentence focus and the underspecified causal structure associated with the "came in contact" verb phrase.

Assuming that stronger sentence-picture match effects are the result of stronger simulations of state change, it is somewhat surprising that sentences with a weaker causal structure elicited more robust simulations. However, this finding is largely consistent with other work examining the role of ambiguity in sentence processing. For example, Mason and Just (2007) used an event-related fMRI paradigm to measure cortical activation when reading sentences featuring a key word with a dominant (unambiguous) meaning (e.g., *ball*) versus a more ambiguous word with multiple, equally likely meanings (e.g., *pitcher*). They found that sentences with ambiguous words elicited broader bilateral hemisphere activation, suggesting that multiple interpretations are simultaneously active during comprehension. Interestingly, these activation patterns were further modulated by individual reading span, where those with higher reading spans were more likely to maintain access to multiple interpretations, such that the unintended dominant meaning required greater suppression.

The latter finding may be particularly informative when considering the results from the norming studies (Appendices B and C). Specifically, we found that the null-subject form of Portuguese led to more salient representations of the target object. We also found that constructions with a limited number of explicit nouns (i.e., those used in Experiment 4a) were often associated with more salient target object representations than other structures with an added pronominal subject. In line with past work (e.g., Gordon et al., 2004; see also Frank & Vigliocco, 2011), we suggest that increasing the number of nouns in the comprehended sentence may have diluted the saliency of *all* described entities, thereby minimizing facilitative effects. This interpretation is also consistent with cross-linguistic effects in the present work, given that the null subject form of Portuguese reduces the number of explicit entities. Taken together, our findings suggest that syntactic constraints do indeed shape simulation processes during comprehension, but that these constraints reflect the joint influence of causal structure, general complexity, and noun order.

Critically, the ordering of nouns is inherently tied to the syntax of the specific language. The present experiments compared English and European Portuguese. Although these languages differ in their subject constructions, both languages follow a Subject-Verb-Object (SVO) order. To date, there is only limited work examining the sentence-picture match effect in languages with other word orders. For example, Sato et al. (2013) examined the representation of object shape in Japanese speakers, for whom the verb-final form (Subject-Object-Verb, SOV) enabled an investigation of the temporal aspects of state change representation. Similar to English speakers, Japanese speakers serially integrated semantic information during reading, and their representations of the described objects were rapidly and appropriately modified after reading a final verb implying state change. However, recall that in Experiment 2, we failed to conceptually replicate the verb phrase manipulation used in Kang et al. (2020a), despite the fact that both studies were conducted in English. Thus, while syntactic differences across languages likely impact mental simulation during sentence comprehension, language alone cannot fully account for the inconsistent sentence-picture match effects observed across experiments.

Future directions and conclusion

Additional work could help substantiate our interpretation of these findings. It would be particularly informative to investigate forms of state change beyond squashing and slicing to better understand the generalizability of the sentence-picture match effect, as well as explore other implicit target object properties. In addition, future work could manipulate sentence focus with these different forms of object state change to better understand how specific verb affordances interact with object salience. Regarding cross-language differences, replicating Experiments 4a and 4b in Portuguese would be especially useful for clarifying whether the findings are limited to English speakers. Lastly, the role of sentence focus could also be explored in languages that follow a SOV (rather than SVO) structure, such as Korean and Japanese.

In conclusion, the present findings challenge the generality of the sentence-picture match effect for object state changes. Our findings suggest that this effect depends on the type of state change, the specific object property used to imply state change, inferred causality, and language-specific syntax, the latter of which modifies which objects are mentally "in focus" during comprehension. Moreover, our findings suggest that the activation—or relative saliency—of a given object state during comprehension may depend on surface-level properties of the sentence. The significant match effect observed in Experiment 4a also stands in contrast to recent failures to replicate the ACE (Morey et al., 2022; Papesh, 2015), which may also be sensitive to sentence focus and syntactic constraints (Jin et al., 2024). Clearly, additional work is needed to determine whether such constraints play a key role not only in simulations of object state changes, but also in the embodiment of described motor movements in language comprehension.

CRediT authorship contribution statement

Channing E. Hambric: Writing – review & editing, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Jon Sides:** Validation, Software, Investigation, Formal analysis, Data curation. **Kevin J. Holmes:** Writing – review & editing, Methodology, Funding acquisition,

Conceptualization. **Abhilasha A. Kumar:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Data and Appendices

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2026.104764.

Data availability

Link is provided in text

References

Alonso-Ovalle, L., Fernandez-Solera, S., Frazier, L., & Clifton, C. (2002). Null vs. overt pronouns and the topic focus articulation in Spanish. *Italian Journal of Linguistics, 14*, 151–170.

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

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

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Bolker, M. B. (2015). Package 'lme4'. *Convergence, 12*(1), 2.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*(2), 193–209.

Briner, S. W., Virtue, S. M., & Schutzenhofer, M. C. (2014). Hemispheric processing of mental representations during text comprehension: Evidence for inhibition of inconsistent shape information. *Neuropsychologia, 61*, 96–104.

Cognola, F., Casalicchio, J., et al. (2018). On the null-subject phenomenon. *Null subjects in Generative Grammar. a Synchronic and Diachronic Perspective*, 1–28.

Connell, L. (2005). Colour and stability in embodied representations. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 27, No. 27).

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

de Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*, 1–12.

del Río, D., Maestú, F., López-Higes, R., Moratti, S., Gutiérrez, R., Maestú, C., & del-Pozo, F. (2011). Conflict and cognitive control during sentence comprehension: Recruitment of a frontal network during the processing of Spanish object-first sentences. *Neuropsychologia, 49*(3), 382–391.

Flusberg, S. F., Holmes, K. J., Thibodeau, P. H., Nabi, R. L., & Matlock, T. (2024). The psychology of framing: How everyday language shapes the way we think, feel, and act. *Psychological Science in the Public Interest, 25*(3), 105–161.

Frank, S. L., & Vigliocco, G. (2011). Sentence comprehension as mental simulation: An information-theoretic perspective. *Information, 2*(4), 672–696.

Frascarelli, M. (2007). Subjects, topics and the interpretation of referential pro: An interface approach to the linking of (null) pronouns. *Natural Language & Linguistic Theory, 25*, 691–734.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558–565.

Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *Cognitive Systems Research, 1*(1), 19–33.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*(1), 97–114.

Greco, A. (2021). Spatial and motor aspects in the "Action-Sentence Compatibility Effect. *Frontiers in Psychology, 12*, 647899.

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.

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

Hoeben Mannaert, L. N., Dijkstra, K., & Zwaan, R. A. (2019). How are mental simulations updated across sentences? *Memory & Cognition, 47*(6), 1201–1214.

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

Horchak, O. V., & Garrido, M. V. (2024). Language comprehenders are sensitive to multiple states of semantically similar objects. *Journal of Memory and Language, 135*, Article 104478.

Jin, H., Zhou, G., & Li, X. (2024). The influence of sentence focus on mental simulation: A possible cause of ACE instability. *Memory & Cognition, 52*(6), 1368–1386.

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

Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., & Altmann, G. T. M. (2020a). The influence of state change on object representations in language comprehension. *Memory Cognition, 48*(3), 390–399.

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

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language, 43*(3), 508–529.

Kaschak, M. P., & Borreggine, K. L. (2008). Temporal dynamics of the action–sentence compatibility effect. *Quarterly Journal of Experimental Psychology, 61*(6), 883–895.

Krass, K. L., Hoang, J. S., Joergensen, G. H., & Altmann, G. T. (2025). Anticipatory eye movements revisited: From affordances and actions to consequences and states. *Brain Research, 149742*.

Lee, S. H. Y., & Kaiser, E. (2021). Does hitting the window break it?: Investigating effects of discourse-level and verb-level information in guiding object state representations. *Language, Cognition and Neuroscience, 36*(8), 921–940.

Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research, 1146*, 115–127.

Morey, R. D., Kaschak, M. P., Dıez-Alamo, M., et al. (2022). A pre-registered, multi-lab non-replication of the action sentence compatibility effect (ace). *Psychonomic Bulletin & Review, 29*(2), 613–626.

Papesh, M. H. (2015). Just out of reach: On the reliability of the action-sentence compatibility effect. *Journal of Experimental Psychology: General, 144*(6), e116–e141.

Prystauka, Y., Wing, E., & Altmann, G. (2024). Investigating the interplay between morphosyntax and event comprehension from the perspective of intersecting object histories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 50*(6), 985–1012.

R Core Team. (2024). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics, 2*(9), 437–442.

Sato, M., Schafer, A. J., & Bergen, B. K. (2013). One word at a time: Mental representations of object shape change incrementally during sentence processing. *Language and Cognition, 5*(4), 345–373.

Santos, A. L. (2005). Getting in focus: The role of the nsr in children's interpretation of sentences with focused preverbal material. In *Language acquisition and development: Proceedings of gala* (pp. 487–498).

Serratrice, L. (2008). Null and overt subjects at the syntax discourse interface: Evidence from monolingual and bilingual acquisition. *LOT Occasional Series, 8*, 181–200.

Soares, A. P., Machado, J., Costa, A., Iriarte, Á., Simões, A., de Almeida, J. J., & Perea, M. (2015). On the advantages of word frequency and contextual diversity measures extracted from subtitles: The case of Portuguese. *Quarterly Journal of Experimental Psychology, 68*(4), 680–696.

Soares, E. C., Miller, P. H., & Hemforth, B. (2019). The effect of verbal agreement marking on the use of null and overt subjects: A quantitative study of first person singular in Brazilian Portuguese. *Forum Linguistico, 16*(1), 3579–3600.

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

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2), 153–156.

Taylor, J. E., Beith, A., & Sereno, S. C. (2020). LexOPS: An R package and user interface for the controlled generation of word stimuli. *Behavior Research Methods, 52*, 2372–2382.

Winter, A., Dudschig, C., Miller, J., Ulrich, R., & Kaup, B. (2022). The action-sentence compatibility effect (ACE): Meta-analysis of a benchmark finding for embodiment. *Acta Psychologica, 230*, Article 103712.

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

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

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