Memory & Cognition (2020) 48:390–399
https://doi.org/10.3758/s13421-019-00977-7

The influence of state change on object representations in language comprehension

Xin Kang[1,2] · Anita Eerland[3] · Gitte H. Joergensen[4,5] · Rolf A. Zwaan[6] · Gerry T. M. Altmann[4,5]

Published online: 17 October 2019

Abstract
To understand language people form mental representations of described situations. Linguistic cues are known to influence these representations. In the present study, participants were asked to verify whether the object presented in a picture was mentioned in the preceding words. Crucially, the picture either showed an intact original state or a modified state of an object. Our results showed that the end state of the target object influenced verification responses. When no linguistic context was provided, participants responded faster to the original state of the object compared to the changed state (Experiment 1). However, when linguistic context was provided, participants responded faster to the modified state when it matched, rather than mismatched, the expected outcome of the described event (Experiment 2 and Experiment 3). Interestingly, as for the original state, the match/mismatch effects were only revealed after reading the past tense (Experiment 2) sentences but not the future-tense sentences (Experiment 3). Our findings highlight the need to take account of the dynamics of event representation in language comprehension that captures the interplay between general semantic knowledge about objects and the episodic knowledge introduced by the sentential context.

Keywords Object state · Mental representation · Language comprehension · Tense · Picture verification

Introduction

One of the dominant features of human language communication is to intentionally talk about absent entities in the past and future that are not immediately spatially or temporally near the speaker and the listener (Cuccio & Carapezza, 2015; Morford & Goldin-Meadow, 1997). This ability to use displaced reference may be unique to humans (Hockett, 1960; Liszkowski et al., 2009) and is linked to the social-cognitive skills of humans and grammatical system of languages (Bergen & Chang, 2005). For example, auxiliary verbs such as "*will*" and "*were*" in English are used to indicate whether an event occurred before or after the moment of speaking or events under discussion. Altmann and Kamide (2007) revealed that visual attention can be directed differently to the visual scene depending on whether the sentences were presented in past-tense or the future-tense conditions. Participants launched more anticipatory eye movements towards a full glass of beer in the future tense (e.g., *The man will drink...*) than in the past tense (e.g., *The man drank...*). Bergen and Wheeler (2010) showed that when we process language that describes perceivable scenes or performable actions, the activation of perceptual and motor systems was influenced by the grammatical markers of tense. For example, when hand motion was described in progressive sentences (e.g., *John is closing the drawer*), there was a facilitation effect for manual action in the same direction of participants, but no such effect was found when the sentences were in perfect tense (e.g., *John has closed the drawer*).

These findings are in line with theories of mental/situation models (Johnson-Laird, 1983; van Dijk & Kintsch, 1983) and

✉ Xin Kang
xin.kang@cuhk.edu.hk

Anita Eerland
a.eerland@uu.nl

1 Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong SAR, China

2 Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong SAR, China

3 Department of Languages, Literature and Communication, Utrecht University, Utrecht, The Netherlands

4 Department of Psychological Sciences, University of Connecticut, Storrs, United States

5 Institute for the Brain and Cognitive Sciences, University of Connecticut, Storrs, United States

6 Department of Psychology, Education, and Child Sciences, Erasmus University Rotterdam, Rotterdam, The Netherlands

perceptual-symbol theories of cognition (Barsalou, 1999) that understanding language involves the construction of a mental situation as a "simulation" of real-world experiences in the spatiotemporal framework, though to what degree perceptual systems are involved in the organization of object knowledge in the brain is still under debate (see Mahon & Caramazza, 2011). According to these theories, concepts of objects are perceptual symbols that arise during perceptual and motor experiences, which can later activate previous experiences and the relevant neural systems. For example, if we think about "*eating an apple,*" we may activate the neural systems of vision, action, touch, taste, and smell that are engaged in our previous experiences. This simulation, or mental models, may include visual information of a red and round object, and sensorimotor information of eating juicy and crunchy pieces.

Using the picture verification paradigm, Stanfield and Zwaan (2001) asked participants to read sentences like "*The carpenter pounded the nail into the wall,*" and to verify whether an object displayed on a picture (e.g., a nail) was mentioned in the sentence. Critically, the object in the picture either matched or mismatched the implied orientation. Although the object's orientation was irrelevant to the task, participants reacted faster to the pictured object (e.g., a horizontally oriented nail) that was compatible with its implied orientation as described in the sentence ("*The carpenter hammered the nail into the wall*") than the incompatible orientation ("*The carpenter hammered the nail into the floor*"). Such match/mismatch effects between linguistic information and visual presentation were found when different properties of objects were manipulated, including orientation (Stanfield & Zwaan, 2001; Wassenburg & Zwaan, 2010), shape (Huettig & Altmann, 2007; Yee, Huffstetler, & Thompson-Schill, 2011; Zwaan, Stanfield, & Yaxley 2002), motion direction (Zwaan, Madden, Yaxley, & Aveyard, 2004), size (de Koning, Wassenburg, Bos, & van der Schoot, 2017), color (Connell, 2007; Hoeben-Mannaert, Dijkstra, & Zwaan, 2017; Huettig & Altmann, 2011; Zwaan & Pecher, 2012), visibility (Yaxley & Zwaan, 2007), distance (Winter & Bergen, 2012), and numerical congruence (Patson, 2016; Šetić & Domijan, 2017).

However, in these studies event models are not established around the target objects alone but draw information from the surrounding environment such as location (e.g., a nail into the floor/wall), other objects (e.g., wine – wine glass), and time (e.g., an hour later vs. a month later). Theories of event models have recognized that events can be encoded across multiple dimensions (Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998; Zwaan, 2016), including location (e.g., Glenberg, Meyer, & Lindem, 1987; Kukona, Altmann, & Kamide, 2014; Radvansky, 2005; Radvansky & Copeland, 2006; Radvansky & Copeland, 2010), time (Radvansky, Zwaan, Federico, & Franklin, 1998; Speer & Zacks, 2005; Zwaan, 1996), goals, and agents. We construct, update, and retrieve the situation models based on these dimensions. When a change occurs in any dimension, we update our mental representations so as to integrate the most recent information and deactivate irrelevant information. It is not always the case that we have to encode events in association with other objects. Linguistic information, such as the tense of sentences, can also be used as a cue for "time shift" (e.g., Altmann & Kamide, 2007; Ferretti, Kutas, & McRae, 2007; Madden & Zwaan, 2003). Besides, previous studies have not clarified whether the object-state has been tracked, maintained, and updated in the event models. For example, an object may go through changes due to an external action (e.g., *The chef chopped the onion*). An onion would look different before and after it is chopped. In this case, the end state of the onion can be distinguished from its initial state and intermediate states from the features of the onion itself. Do we also encode the conflicting states of the onion in our event models?

So far, there is limited empirical evidence of the encoding of object state-change in language comprehension (but see Altmann, 2017; Hindy, Altmann, Kalenik, & Thompson-Schill, 2012; Solomon, Hindy, Altmann, & Thompson-Schill, 2015). Hindy et al. (2012) revealed that reading sentences that described a change of state provides a challenge to our cognitive system; multiple representations of the object in different states may be activated and we have to choose the situationally appropriate one. Solomon et al. (2015) further revealed that competition between object states (e.g., an onion in its original state or its subsequent chopped state) is only revealed when the states are associated with the same object, but not a different version of that object (e.g., one onion in its original state, and another onion in a chopped state). Thus, it is likely that when a change of state event occurs, the object is linked to multiple "states" of itself across time – before and after this change. Therefore, an object in the modified state has its own "history" that includes its association with its prior original self and with changes to its states across time. Altmann and Ekves (2019) further proposed the "events as intersecting object histories" (IOH) model that encoding events (whether we directly experience them or learn about them through language) involves constructing dynamic representations of intersecting object histories. If mental simulations of situations are an integral part of understanding language, can we expect match/mismatch effects after the object experiences a change of state?

In the present study, we aimed to explore whether object state-change influences the speed of picture verification. We expected to find quicker response times when the object representation matched the picture probe compared to when it mismatched the probe. Experiment 1 was intended to establish baseline responses to probe pictures that showed conflicting states of target objects. In Experiment 2, we manipulated object state-change by using two different verbs – one indicating a minimal/no change and the other a substantial change. An example is *The woman chose/dropped the ice cream.* The task for participants was to verify whether the probe picture that appeared afterwards was mentioned in the sentence they just read. Our hypothesis was that

despite the irrelevance of object states to the verification task, the object states that are activated in language would influence the responses to probe pictures. Thus, we predicted that participants would react faster to a probe picture when it matched the described state of the target object than when it mismatched the object state. Experiment 3 further explored whether the tense of sentences would further mediate the activation of object states in language comprehension when the sentences were in future tense.

Data were collected via Amazon's Mechanical Turk (MTurk, http://www.mturk.com) using the sentence-picture verification paradigm following Zwaan and Pecher (2012). All the materials and raw data for our study can be found on the Open Science Framework (OSF) (https://osf.io/cvrm3/). The key independent variable was whether the state of the picture was compatible or incompatible with the original state and modified state of object described in the sentences. On each trial, participants read a word (e.g., *ice cream*) or a sentence (e.g., *The woman dropped the ice cream*) and then indicated whether a subsequent picture showing an ice cream was mentioned in the text. Visual probes were committed to a particular shape of the target object and thus allowed us to assess the activation levels of different forms of the same entity. Accuracy and reaction times to the probe pictures were recorded. Response times (RTs) were calculated over correct trials only and RTs that were shorter than 300 ms or longer than 3,000 ms were excluded. The linear mixed-effects models (LMMs) using the lme 4 package (Bates et al., 2015; Baayen, Davidson, & Bates, 2008) of R (R CoreTeam, 2016) were used for statistical analysis. The lsmeans package (Lenth, 2016) was used to conduct post hoc comparisons for significant interaction effects with Tukey adjustments of *p*-values. Table 1 summarizes fixed effects of LMMs in all three experiments.

Experiment 1

We conducted our first experiment using the word-picture verification paradigm to identify the baseline responses towards our picture stimuli. This allowed us to determine whether one state would be responded to differently from the other state when only the object's name was mentioned. We selected 32 high-frequency object names (e.g., *ice cream, banana, rope, candle*) and paired each name with one of two pictures of the object – one showing the object in the original state (e.g., *an upright ice cream*) and one showing the object in a modified state (e.g., *a dropped ice cream*).

Table 1 Fixed effects estimated with linear mixed models in all experiments

	Fixed effects	Coefficient	SE	t	*p*
Experiment 1	Picture type	117	18	6.53	<.001
Experiment 2	Picture type	153	18	8.58	<.001
	Event type	83	18	4.69	.615
	Interaction	153	25	6.12	<.001
Experiment 3	Picture type	175	49	5.18	<.001
	Event type	151	34	4.48	.001
	Interaction	145	47	3.06	.002

Method

Participants We recruited 118 participants (54 female, mean age 36.19 years, range 19–64) through MTurk. All participants were residents of the USA and received US$1.50 for their participation, which lasted approximately 15 min. One participant was excluded for reporting a non-English native language. With the exclusion of this participant our sample included 117 native English speakers.

Materials Each participant saw one of two lists that counterbalanced items and conditions of experimental trials. Crucial to the goal of Experiment 1, the object name (e.g., *ice cream*) in the experimental trials could be followed by a picture showing this object either in its original state (e.g., an upright ice cream) or in its modified state (e.g., a dropped ice cream) that are caused by external forces but not an internal action (e.g., blooming flowers) (see White, 1991, for the distinction between external/internal causal attribution). Figure 1 illustrates two example pairs of probe pictures.

We created four practice trials (including two "yes" responses and two "no" responses), 32 experimental trials requiring "yes" responses (e.g., the object's name "*ice cream*" followed by a picture of an *ice cream*), and 32 fillers requiring "no" responses (e.g., the object's name "*box*" followed by a picture of a *ball*). All pictures were from a commercial clipart website and were edited to best match the intended states of the object. The pictures were resized to a maximum of 3 in. height and 3 in. width.

Procedure The experiment was presented online in the Qualtrics survey research suite (http://www.qualtrics.com). Each trial started with the presentation of a left justified and vertically centered fixation cross on the computer screen for 1,000 ms, immediately followed by an object name (e.g., ice cream), centered at the same location as the fixation cross. Participants pressed the spacebar when they had read and understood the object name. After the keypress, the object name (e.g., ice cream) was replaced by a fixation-cross that appeared for 500 ms, and then immediately followed by a picture (e.g., an upright ice cream). Participants were instructed to indicate as fast and accurately as possible whether the object displayed in the picture matched the object name they just read (yes/no) by pressing a button on the keyboard (m-key/c-key, respectively). All experimental trials required a "yes" response, whereas all filler trials required a "no" response. The next trial started 500 ms after the

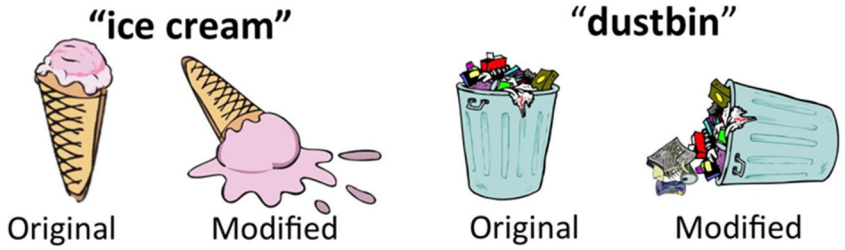

Fig. 1 Two example pairs of probe pictures used in Experiments 1–3. The original state depicts a canonical or prototypical form of the object, while the modified state was usually caused by an external action (e.g., *drop*)

response was given. Experimental and filler trials were presented in random order.

Results and discussion

We estimated the fixed effects of Picture type (Original vs. Modified) with subjects and items as random effects in the first model.

$$model1 < - lmer(RT \sim Picture + (1|Subject) + (1|Item))$$

The second model is a random-intercept-and-slopes model without fixed effects.

$$model2 < - lmer(RT \sim 1 + (1|Subject) + (1|Item))$$

The models were fit by restricted maximum likelihood (REML). To assess the goodness of fit, we compared the models using the χ^2-distributed likelihood ratio and its associated *p*-value. The model with a smaller Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) was considered as a better fit. The results showed that participants responded significantly faster to the original state, $\chi 2\ (2) = 41.177$, $p<.001$, suggesting the original state (LSMEANS: 893 ± 50 ms) seems to have an advantage in response times compared to the modified state (LSMEANS: 1009 ± 50 ms) (see Fig. 2)

Experiment 2

In Experiment 2, we examined whether the linguistic context may modulate picture verification responses.

Method

Participants We recruited 227 participants (100 female, mean age 35.07 years, range 18–65) through MTurk. All participants were residents of the USA and received US$2.00 for their participation, which lasted approximately 25 min. Thirty-one participants indicated a language other than English as their native language. With the exclusion of these participants, our sample included 196 native English speakers.

Materials and procedure We created four lists of stimuli (two types of events x two types of pictures) to counterbalance items and conditions. The procedure of Experiment 2 was identical to that of Experiment 1. Participants read 32 experimental sentences that either described a substantial change of an object's state (e.g., "*The woman dropped the ice cream*") or minimal/no change (e.g., "*The woman chose the ice cream*"). After reading the sentences, participants pressed the SPACE bar and a picture probe appeared showing either the original state or the modified state of that object. In addition to these 32 experimental items that required "yes" responses, 32 filler items that required "no" responses were added (e.g., "*The man kicked the ball*"*;* a picture of a box). Items were presented in random order.

Results and discussion

We adopted the same statistical analysis procedure as Experiment 1. The fixed effects were Picture type (Original vs. Modified) and Event type (Substantial change vs. Minimal change). Random effects included subjects and items. The goodness of fit was assessed by comparing the AIC and BIC values and χ^2-distributed likelihood ratio and its associated *p*-value. We found a significant fixed effect of Picture type, $\chi^2\ (1) = 35.85$, $p < .001$ with faster reaction times to the original state than to the modified state. No fixed effect of the Event type was found. Nonetheless, there was a significant interaction between Picture type and Event type, $\chi^2\ (1) = 37.322$, $p < .001$. Post hoc comparisons indicated that the original state was verified faster when the sentence implied an event involving a minimal change of state (LSMEANS: 1,091 ± 33 ms) than a substantial change of state (LSMEANS: 1,161 ± 34 ms; $p < .001$), while the modified state was verified faster when the sentence implied an event involving a substantial change of state (LSMEANS: 1,161 ± 33 ms) than a minimal change of state (LSMEANS: 1,244 ± 34 ms, $p < .001$) (see Fig. 3).

Therefore, it seems that a match advantage of the original state and modified state was found when they were presented after the condition that indicated the appropriate end state of the object. In sum, the results of Experiment 2 suggest that the linguistic context has an impact on the activation of object-state representations. The original state of an object was

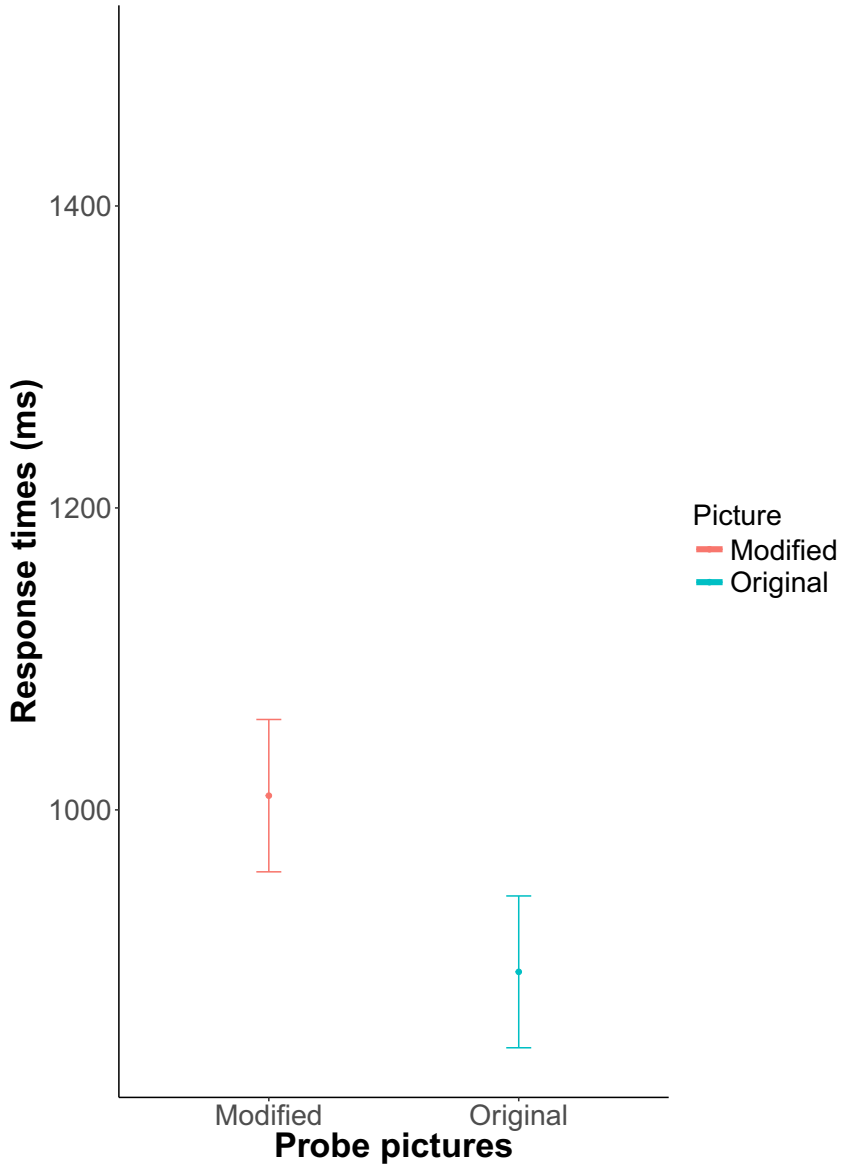

Fig. 2 Mean response times of the probe pictures after reading the object name in Experiment 1. Data are shown as LSmean ± SE. The y-axis shows the response times to probe pictures in milliseconds (ms)

verified faster when the sentences described a minimal change of state, but when a substantial change of state was described the modified state was verified faster. Despite the fact that the original state was verified faster than the modified state in Experiment 1, the modified state gained a match advantage when it was the expected end state of the event.

Our findings are consistent with previous research showing that contextually appropriate perceptual information about described objects is activated in language comprehension (e.g., Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002; see also Hoeben-Mannaert et al., 2017), and the evidence on the competition between multiple object states in language comprehension (e.g., Hindy et al., 2012). Importantly, this experiment demonstrates that the internal structure of a narrated sequence of events can be mapped with the representation of objects. Previous studies have often specified the location (e.g., *on the wall/floor*) or the time (e.g., *after 1 day/1 year*) as the cues for such event sequences. Our findings show that these explicit cues may not be necessary as the states of the object can be described by using the verbs (e.g., *drop*), which will also trigger the activation of the corresponding object state that is the appropriate end state of the event.

Experiment 3

The first two experiments showed that (1) without any linguistic context, people mentally represent the original state of an object, and (2) the modified state has a match advantage when linguistic context indicates a change compared to no change. When we read the past tense version of a sentence (e.g., *The woman dropped the ice cream*), relative to the time of the hearer, the event has happened in the past, meaning that the ice cream is already in its dropped state.

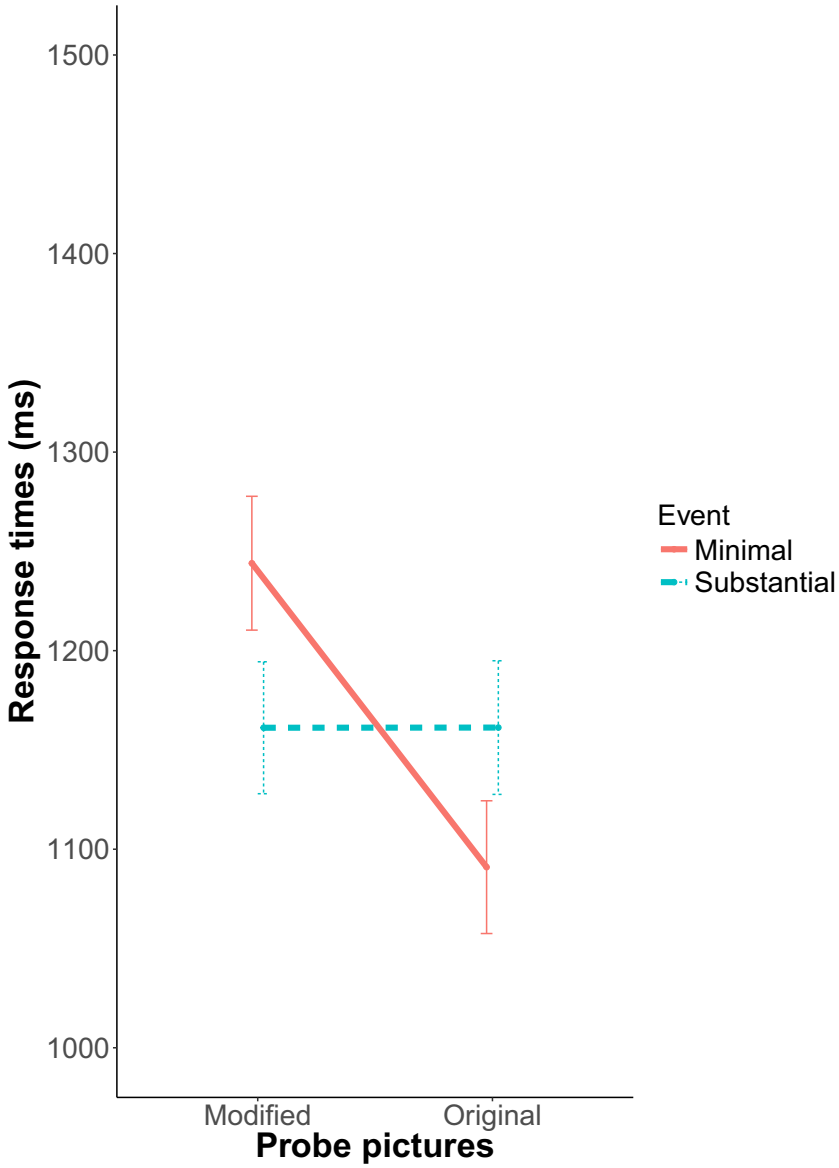

Fig. 3 Mean response times of the probe pictures after reading past-tense sentences such as "*The woman dropped/chose the ice cream*" in Experiment 2. Data are shown as LSmean ± SE. The y-axis shows the response times to probe pictures in milliseconds (ms)

However, if a sentence is in the future tense (e.g., *The woman will drop the ice cream*), relative to the time of the hearer, the ice cream is original, and although a change in state is described, if the hearer were to act on the ice cream, it would be in the original state. Thus, theoretically we would inhibit the activation of the changed state of the ice cream and keep the original state as being more accessible. When the participant-centric current state of the world entails an original ice-cream, but a future state of the world entails the dropped ice cream – will the representation associated with the current state be the more accessible, or will the representation with the as-yet un-encountered future state be more accessible?

In Experiment 3, we aimed to investigate whether the tense of a sentence modulates the activation of the most prominent states of the object. Previous studies have shown that grammatical tense may play a role in constructing mental representations. In Bergen and Wheeler (2010), participants read sentences that were in the present progressive tense (Experiment 1: e.g., *Carol is taking off/putting on her glasses*) or in the present perfect tense (Experiment 2: e.g., *Carol has taken off/put on her glasses*) and decided if the described action required movement of the hand toward or away from the body. They found an action-sentence congruency effect for the progressive sentences in Experiment 1 but not the perfect sentences in Experiment 2, arguing that the actions in Experiment 2 were already completed and hence

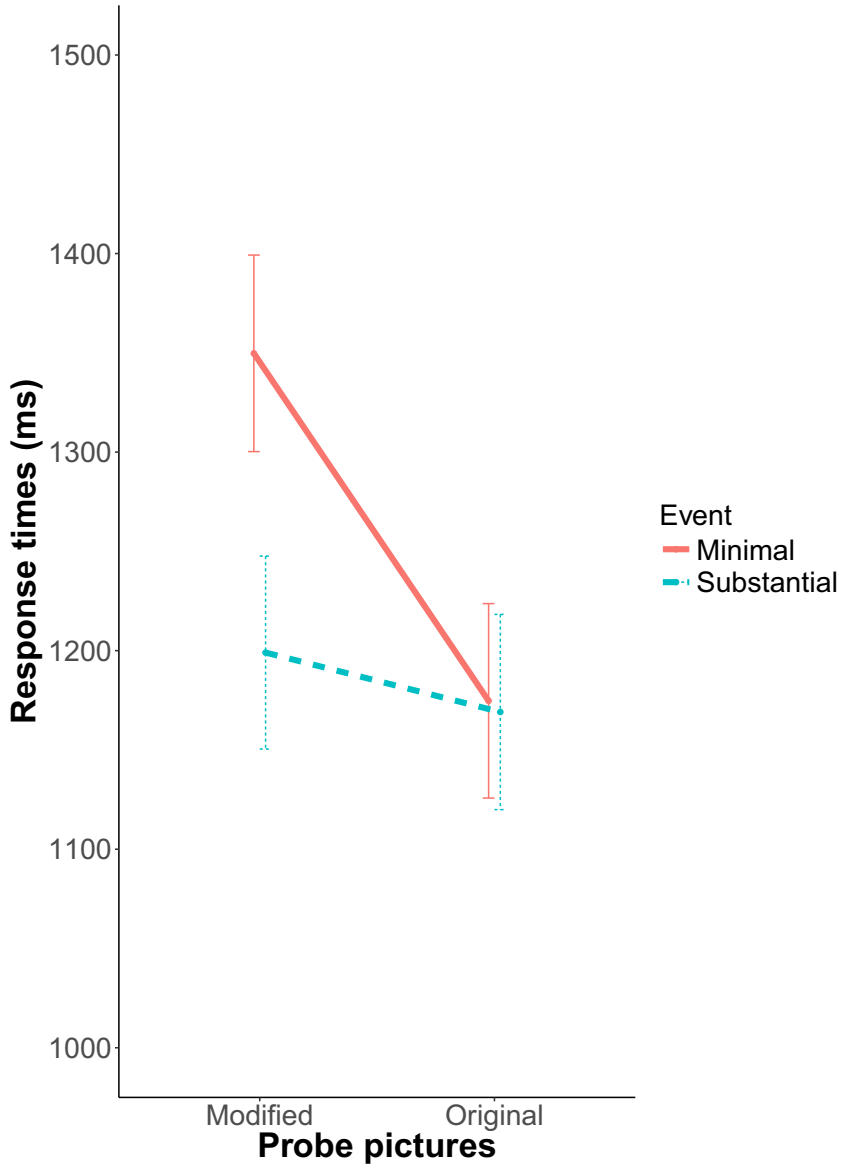

Fig. 4 Mean response times of the probe pictures after reading the future-tense sentences such as "*The woman will drop/choose the ice cream*" in Experiment 3. Data are shown as LSmean ± SE. The y-axis shows the response times to probe pictures in milliseconds (ms)

required no simulation of the action. If this is the case, consider "*The woman will drop the ice cream,*" which might establish two conflicting states of the object (i.e., the current original state and the future modified state). Will there be a match advantage for the original state given that a possible change of state has not happened yet?

Method

Participants We recruited 211 participants (104 female, mean age 33.87 years, range 18–69) through MTurk. All participants were residents of the USA and received US$2.00 for their participation, which lasted approximately 25 min. Six participants indicated a language other than English as their native language. With the exclusion of these participants, our sample included 205 native English speakers.

Materials and procedure Experiment 3 was identical to Experiment 2 with the exception that all single sentences in this experiment used future tense rather than past tense.

Results and discussion

We followed the same statistical procedure as in Experiment 2. The results suggested that there was a fixed effect of Picture type, $\chi 2\ (1) = 18.18$, $p < .001$ that the original pictures were responded to faster than the modified pictures. There was a fixed effect of Event type, $\chi 2\ (1) = 10.65$, $p = .001$ that the picture following a

substantial change was verified faster than a minimal change. Importantly, there was a significant interaction between Picture type and Event type, $\chi 2$ (1) = 9.36, p =.002. Post hoc comparisons showed that there was no significant difference in verification time between a minimal change (LSMEANS: 1,175 ± 49 ms) and a substantial change (LSMEANS: 1,169 ± 49 ms, p = .998) when the probe picture indicated an original state of the object. However, the modified state was verified significantly faster, when the sentence indicated a substantial change of state event (LSMEANS: 1,199 ± 49 ms) than a minimal change of state event (LSMEANS: 1,349 ± 49 ms, $p < .001$) (see Fig. 4).

These results are partly consistent with the findings of Experiment 2 in that there was a match advantage for the modified state in the change situation compared to the no-change situation. That is, when a picture of a "dropped" ice cream is presented to participants, they might associate the picture with the "*drop*" action, but not necessarily with the "*choose*" action. However, as opposed to Experiment 2, the original state did not show any match advantage in the minimal change condition compared with the substantial change condition. Previous empirical studies have shown facilitatory effects between action and affordances of objects (e.g., Symes et al., 2007). Our findings may be accounted by the equal affordance of the original state for a minimal change and a substantial change in the future tense. By comparison, in the past tense, the original state matched the end state of the minimal change, but it did not afford further substantial changes and did not match the consequences, leading to the match/mismatch effect.

General discussion

In this study, we reported findings of three experiments that explored the activation of objects' mental representations in language comprehension. More specifically, we investigated the influence of object-state (original vs. modified) that was manipulated by degree of change of the event (a minimal change vs. a substantial change) and grammatical tense (past tense vs. future tense) on picture verification responses. Our study showed that the original state was responded to faster than the modified state when only object name was presented (Experiment 1). Nonetheless, when contextual information was provided, the described situation in language modulated the responses. Objects in the modified state were verified more quickly when they was described to experience a substantial change than a minimal change sentences in both past tense (Experiment 2) and future tense (Experiment 3). However, objects in their original state were only verified more quickly when they were described to experience a minimal change than a substantial change condition in past tense sentences (Experiment 2) but not in future-tense sentences. Our results suggested that the activation of the contextually appropriate object representation was modulated by the degree of change. Our findings also indicated that there was a close link between the consequences of the action and the grammatical tenses of sentences. This tight coupling between action and knowledge of objects supports the IOH model (Altmann & Ekves, 2019) that language comprehension involves activating situated object states before and after object state-change.

One limitation of this study is that we measured the activation of object representation at the end of sentence reading, which may only be able to capture part of the activation processing. Another potential confound was that the degree of change was manipulated by using two different verbs. Thus, the effects that we observed might be driven by the semantic associations between the actions and the perceptual properties of the objects (e.g., Bach, Nicholson, & Hudson, 2014). For example, the dropped ice cream could be associated with the "*drop*" action but not the "*choose*" action. Without any simulation of the motor or perceptual properties of the situation, one may even establish the link between the object and the consequences of the action.

In conclusion, our experiments demonstrate that perceptual properties of objects can be activated in language comprehension and modulated by the content of the linguistic input. The interplay between general semantic knowledge about objects and the episodic knowledge introduced by the sentential context is captured by dynamics of event representation in language comprehension.

Data accessibility statement The materials and data for all the experiments can be found on the Open Science Framework (https://osf.io/cvrm3/).

Author contributions Contributed to conception and design: XK, AE, GHJ, RAZ, GTMA
Contributed to acquisition of data: XK, GHJ, GTMA
Contributed to analysis and interpretation of data: XK, AE, GHJ, RAZ, GTMA
Drafted and/or revised the article: XK, AE, GHJ, RAZ, GTMA
Approved the submitted version for publication: XK, AE, GHJ, RAZ, GTMA

Funding information This research was supported by an Overseas Research Scholarship from the University of York awarded to Xin Kang.

Compliance with ethical standards

Competing interests The authors declare that they do not have any competing interests.

References

Altmann, G. T. M. (2017). Abstraction and generalization in statistical learning: implications for the relationship between semantic types and episodic tokens. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711).

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

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57,* 502–518.

Bach, P., Nicholson, T., & Hudson, M. (2014). The affordance-matching hypothesis: how objects guide action understanding and prediction. Frontiers in Human Neuroscience, 8, 254.

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences, 22(4),* 577–609; discussion 610–660.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, Articles*, *67*(1), 1–48.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bergen, B., & Chang, N. (2005). Embodied construction grammar in simulation-based language understanding. In J.-O. Ostman & M. Fried (Eds.), *Construction Grammars: Cognitive Grounding and Theoretical Extensions* (147–190). Amsterdam: Benjamins

Bergen, B., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain and Language*, *112*(3), 150–158.

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

Cuccio, V., & Carapezza, M. (2015). Is displacement possible without language? Evidence from preverbal infants and chimpanzees. *Philosophical Psychology,* 28(3), 369–386.

de Koning, B. B., Wassenburg, S. I., Bos, L. T. & van der Schoot, M. (2017. Mental simulation of four visual object properties: similarities and differences as assessed by the sentence-picture verification task. *Journal of Cognitive Psychology. 29*, 420-432.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(1), 182–196.

Glenberg, A.M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26,* 69–83.

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

Hockett, Charles F. (1960). The origin of speech, *Scientific American,* 203, 88-111.

Hoeben-Mannaert, L., Dijkstra, K., & Zwaan, R.A. (2017). Is color an integral part of a rich mental simulation? *Memory & Cognition, 45*, 974-982.

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15*(8), 985–1018.

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing "frog": How object surface colour and stored object colour knowledge influence language-mediated overt attention. *The Quarterly Journal of Experimental Psychology, 64*(1), 122–145.

Johnson-Laird, P.N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Kukona, A., Altmann, G. T. M., & Kamide, Y. (2014). Knowing what, where, and when: event comprehension in language processing. *Cognition*, *133*(1), 25–31.

Lenth, R. (2016). Least-Squares Means: The R Package lsmeans. Journal of Statistical Software, Articles, 69(1), 1–33.

Liszkowski, U., Schäfer, M., Carpenter, M., & Tomasello, M. (2009). Prelinguistic infants, but not chimpanzees, communicate about absent entities. *Psychological Science*, *20*(5), 654–660.

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

Morford, J. P., & Goldin-Meadow, S. (1997). From Here and Now to There and Then: The Development of Displaced Reference in Homesign and English. *Child Development*, *68*(3), 420.

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in Cognitive Sciences*, *15*(3), 97–103.

Nikole D. Patson, (2016) Evidence in support of a scalar implicature account of plurality.. Journal of Experimental Psychology: Learning, Memory, and Cognition 42 (7):1140-1153

Bach, P., Nicholson, T., Hudson, M. (2014) The affordance-matching hypothesis: how objects guide action understanding and prediction. Frontiers in Human Neuroscience 8

Radvansky, G. A. (2005). *Situation models, propositions, and the fan effect.* Psychonomic Bulletin & Review, 12(3): 478-483.

Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & Cognition, 34(5),* 1150-1156.

Radvansky, G. A. & Copeland, D. E. (2010). Reading times and the detection of event shift processing. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 36,* 210-216.

R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Radvansky, G. A., Zwaan, R. A., Federico, T., & Franklin, N. (1998). Retrieval from temporally organized situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1224-1237.

Šetić, M., & Domijan, D. (2017). Numerical congruency effect in the sentence-picture verification task. Experimental Psychology, 64(3), 159-169.

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

Speer, N.K., Zacks, J.M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language, 53,* 125-140.

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

Symes, E., Ellis, R., & Tucker, M. (2007). Visual object affordances: object orientation. *Acta Psychologica*, *124*(2), 238–255.

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

Wassenburg, S. I., & Zwaan, R. A. (2010). Readers routinely represent implied object rotation: the role of visual experience. *Quarterly Journal of Experimental Psychology, 63*(9), 1665–1670.

White, P. A. (1991). Ambiguity in the internal/external distinction in causal attribution. *Journal of Experimental Social Psychology*, *27*(3), 259–270.

Winter, B., & Bergen, B. (2012). Language comprehenders represent object distance both visually and auditorily. *Language and Cognition*, *4*(1), 1–16.

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition*, *105*(1), 229–236.

Yee, E., Huffstetler, S., & Thompson-Schill, S. L. (2011). Function follows form: Activation of shape and function features during object identification. *Journal of Experimental Psychology: General, 140*(3), 348.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The Construction of Situation Models in Narrative Comprehension: An Event-Indexing Model. *Psychological Science*, *6*(5), 292–297.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* 22 (5): 1196-1207.

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, *23*(4), 1028–1034.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: dynamic representations in language comprehension. *Cognitive Science, 28*(4), 611–619.

Zwaan, R.A., & Pecher, D. (2012). Revisiting Mental Simulation in Language Comprehension: Six Replication Attempts. *PLoS ONE, 7*, e51382.

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