Dropping Bowling Balls on Tomatoes: Representations of Object State-Changes During Sentence Processing

Oleksandr V. Horchak & Margarida Vaz Garrido

Iscte-Instituto Universitário de Lisboa, Cis-Iscte, Lisboa, Portugal

Authors' notes

Acknowledgments: This research was supported by the Foundation of Science and Technology of Portugal by a grant SFRH/BPD/115533/2016 awarded to the first author. We would like to thank Gerry Altmann and two anonymous reviewers for their valuable suggestions on previous versions of the manuscript.

Data Accessibility Statement: All data, materials, and analysis scripts are openly available at the website of the Open Science Framework (https://osf.io/fd52u/?view_only=5e336b8086f34848af84aa54811700cc)

Author Contributions

Study idea: OH. Design of sentence stimuli: OH, MG. Design of picture stimuli: OH. Acquisition of data: OH. Analysis and interpretation of data: OH, MG. Draft of the manuscript: OH. Critical revision of the manuscript: OH, MG. Approval of the submitted version for publication: MG, OH.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Correspondence should be sent to Oleksandr V. Horchak, ISCTE-IUL, Av. das Forças Armadas, 1649-026, Lisbon, Portugal. E-Mail: Oleksandr.Horchak@iscte-iul.pt

Abstract

Previous research showed that verifying a pictured object mentioned in a preceding sentence takes less time when the pictured object shape is compatible with the described object location or spatial position. In the current work we asked if non-visual information is integrated into the mental model when the target object shape is implied by virtue of a description of a heavy vs. light item being dropped on it. Furthermore, we asked if the canonical target object state continues to play an important role when the context requires the activation of a non-canonical representation. In seven experiments the data provide an affirmative response to both questions. Participants ($N = 766$) first read sentences that implied target object state-changes as a function of the impact caused by differently-weighted items (e.g., “You drop a balloon/a bowling ball on a tomato”) and then verified pictures of “squashable” target objects in either a canonical (e.g., intact tomato) or a non-canonical (e.g., squashed tomato) state. A reaction time advantage was consistently observed when a “non-canonical” target was preceded by a “heavy” (e.g., bowling ball) sentence than a “light” (e.g., balloon) sentence. However, no such advantage was observed when a “canonical” target was preceded by a “light” sentence than a “heavy” sentence. Importantly, this pattern of results remained unchanged regardless of the items used and the verbal tense of the sentence. These data suggest that when changes of state are inferred (i.e., not driven by lexical semantics), both the initial and resultant states are equally accessible.

Keywords: object state, mental representation, language comprehension, weight, perception, action

Introduction

Imagine you are reading a sentence about a boy dropping a balloon on a tomato. If you were asked to picture the described tomato in your mind, what image would you be likely to think of first: a round tomato, or a squashed tomato? Undoubtedly, the former image would come to mind of anyone with some knowledge of how heavy balloons are. But what if an immediately succeeding sentence described a boy dropping a bowling ball on a tomato? Would you picture the initial state of a tomato (round shape), the intermediate state (associated with the perception of a collision of one object with another) of a tomato, or the end state (deformed shape) of a tomato?

The situation model theory (Kintsch & van Dijk, 1978), the mental-model theory (Johnson-Laird, 1983), and the event-indexing-model theory (Radvansky et al., 1998; Zwaan, Langston, et al., 1995; Zwaan, Magliano, et al., 1995; Zwaan & Radvansky, 1998) suggest that our ability to process the abovementioned information is facilitated through the construction of mental representations of entities and events described in a text rather than the structure of the text itself (Zwaan, 1999). In these theories, situation model consists of multiple, hierarchically-represented events, that are related to one another on several dimensions. According to the event-indexing model (e.g., Zwaan & Radvansky, 1998), for example, comprehenders index, encode, and update each event mentioned in the story on at least the following five dimensions: protagonist, motivation, time, space, and causation. Despite a long history of research on the content of such mental representations (e.g., location: Kukona et al., 2014; time: Speer & Zacks, 2005; causation: Gernsbacher, 1990), most theories of event cognition and event representation do not consider the relevance of object-state change for event representation (cf. Altmann & Ekves, 2019). To our knowledge, no study so far has considered the integration of implied visual, action, proprioceptive, and kinesthetic information into the mental model to be able to convincingly state which of the aforementioned object states (i.e., a round tomato, a squashed tomato, etc.) may get activated or deactivated in the mental model. Therefore, over the course of seven experiments we present initial evidence consistent with (1) the idea that non-visual features of the situation (i.e., the weight of an item that falls on a target object) are taken into account when representing target object shape and (2) the proposal that, at least under some circumstances, prototypical target object information, which is initially activated (e.g., a round tomato), cannot be completely overwritten or inhibited even when the content of the linguistic input requires the activation of a different representation (e.g., a squashed tomato).

Previous research

A popular paradigm to reveal the content of mental representations is a sentence-picture verification task in which participants read a sentence and then are shown a pictured object. For instance, Zwaan et al. (2002) asked participants to read sentences like “A ranger saw an eagle in the sky” or “A ranger saw an eagle in a nest” and then judge if a subsequently presented pictured object was mentioned in the sentence. Participants’ responses were faster when the shape of a pictured object (e.g., an eagle with outstretched wings vs. an eagle with folded wings) matched the shape of the object implied by the linguistic description. Similar findings regarding object shape were also reported by Engelen et al. (2011), Pecher et al. (2009), Rommers et al. (2013), and Zwaan and Pecher (2012).

Although there is now a wealth of evidence as to what object properties are activated in mental representations (e.g., Horchak et al., 2014, Horchak & Garrido, 2020), researchers are now increasingly addressing the question of how readers activate such mental representations, including those for object shape. For example, Ferguson et al. (2013) illustrated that contextual uncertainty about the described event influences the content of mental representations. More specifically, they showed that participants were significantly faster to verify a matching picture of the target image (following a delay of 250 ms) after reading a sentence such as “The old lady *knows* that the picnic basket is open” than a sentence such as “The old lady *thinks* that the picnic basket is open”, thus suggesting that in uncertain conditions a construction of a mental representation is a more time-consuming process than in certain conditions. Altmann and Kamide (2009) used eye tracking to investigate the mapping between language input and mental representations of visual scenes. By manipulating the event-related locations of objects, they found that participants landed more fixations on the table at the offset of the word “glass” in “She will pick up the bottle, and pour the wine carefully into the glass” when preceded by a sentence “The woman will put the glass onto the table” than when preceded by a sentence “The woman is too lazy to put the glass onto the table”. This finding thus suggests that during the process of object recognition comprehenders constantly update event-based representations of observed referents. Sato et al. (2013) provided direct empirical evidence for the dynamically updateable event-based representations of object shape by using Japanese language in the picture-verification task. The processing of sentences in Japanese, which has a verb-final order, created an expectation of one object state at the offset of a sentence and a different object state at the end of the sentence (e.g., first reading about a man wearing a kimono and then processing the verb that implies that the kimono has been torn apart). The researchers found that participants’ verification of shape-matching pictures was significantly faster both before (e.g., not damaged kimono) and after the presentation of the critical final verb contradicting the initially expected object state (e.g., damaged kimono), thus pointing to the conclusion that mental representations of object shape get activated both in the middle and at the end of the sentence. Finally, Hoeben et al. (2019) have recently found that the initial object state is quickly revised when the other object state is mentioned. They did so by presenting participants with a set of sentences in which an object was dynamically changing from one shape (e.g., an eagle with outstretched wings in the sky) to another (e.g., an eagle with folded wings in the nest) as a function of location. More specifically, the data revealed that verification times were faster for the most recently implied shape (i.e., an eagle with folded wings), thus suggesting that the end object state was more activated.

What remains unclear, however, from the above findings is whether the initial object state is as rapidly revised when object state-change is contingent on action. Altmann and Ekves (2019) in their “Events as intersecting object histories (IOH)” account argued that representational consequences for the changes of location (as compared to action) are different, given that changes in the surrounding context in which an object is described require encoding of that context. Indeed, when event models are established from the changes in location (e.g., “There is an egg in a fridge” vs. “There is an egg in a skillet”), the transition from an object being intact to it being crushed is occluded (although the object in its crushed state will activate semantic knowledge of the object in general), and hence the comprehenders should be, at the very minimum, less sensitive to the activation of an earlier part of an object’s trajectory. However, this is not the case when an object is described as substantially changing state due to an external action. For a sentence such as “The man *dropped* the glass”, it makes sense to predict that the *resultant* state cannot be divorced from the *original* state, precisely because one needs to know what an initial object state was in order to comprehend that a change in state actually occurred. Such a prediction fits with Altmann's and Ekves' (2019) theoretical account, which predicts the anticipation of goal states given all other possible states.

That there may be a competition between object states in event comprehension is also supported by empirical evidence. Using functional magnetic resonance, Hindy et al. (2012) presented participants with sentences in which an object was described as changing substantially (e.g., “The squirrel will crack the acorn”) compared to changing minimally (e.g., “The squirrel will sniff the acorn”). The researchers found that a neural marker for competition was present in the "crack" case more than in the "sniff" case, and they concluded that for competition to obtain in these cases, multiple states of the acorn had to be co-activated. Furthermore, a subsequent study by Solomon et al. (2015) confirmed that this competition required distinct states of the same acorn.

Most recently, behavioral evidence was provided in support of an idea that language processing involves activating relevant object states both *before* and *after* object state-change. Kang et al. (2019) have conducted a series of picture verification experiments in which participants read a word or a sentence and subsequently saw a picture. The task was to indicate whether the object was mentioned in the word or the sentence. In Experiment 1, the researchers presented participants with object names (e.g., ice cream) that were followed by a picture depicting the object in a normal or a crushed state and found that the intact object state had a substantial advantage in response times (difference more than 100 ms) compared to the crushed object state. In Experiment 2, participants saw the same picture stimuli as in Experiment 1, except that these were now preceded by past-tense sentences describing an action that would leave an object in its original state (e.g., The woman *chose* the ice cream") or an action that would crush the object (e.g., The woman *dropped* the ice cream"). The results now showed that picture verification times were shorter for both the original and modified states of the object whenever the pictured target's state matched the end state implied by the sentence. In Experiment 3 participants saw the same sentences and pictures as in Experiment 2, with an exception that sentences were presented in the future tense (e.g., "The woman *will drop/choose* an ice-cream"). This time the results demonstrated that depictions of deformed objects showed the matching effect in the substantial change ("drop" sentence) condition, but pictures of intact objects did not show the matching effect in the minimal change ("choose" sentence) condition. Finally, in both Experiments 2 and 3 no significant response time advantage was observed for the pictured original object state (i.e., intact ice- cream) relative to the pictured modified object state (i.e., squashed ice-cream) in the substantial change (“drop” sentence) condition. Kang et al. (2019) concluded that the interplay between world knowledge about objects and the grammatical tenses of sentences defines the dynamics of event representation.

Three important conclusions can be drawn from Kang et al.'s (2019) study. First, when the degree of change is manipulated by using two different verbs (e.g., choose vs. drop) with “squashable” objects, the initially activated object information can only be successfully updated when the past tense of the sentence “forces” comprehenders to focus on the completed action. Second, if a sentence is in the future tense (e.g., will drop vs. will choose), the activation of the crushed change of an object is partially inhibited, given that from the participant-centered perspective an original object representation is more accessible at the moment of the action of dropping. Finally, the *before* and *after* states of an object compete during event representation, as evidenced by no response time advantage in the original pictured object state (i.e., intact ice- cream) relative to the modified pictured object state (i.e., squashed ice-cream) in the substantial change (“drop” sentence) condition.

Nonetheless, if event models draw information from visual features of the situation (e.g., locations) and different actions, then it stands to reason that unmentioned, non-visual features of the situation (e.g., when the shape of a target object is implied by virtue of a description of a heavy vs. light second object being dropped on it) should also affect mental representations of described situations. In support of such an idea is a study of Scorolli, Borghi, and Glenberg (2009) showing that such an intrinsic object property as weight is simulated during language comprehension. In this research, participants lifted differently weighted (but visually identical) boxes after reading sentences describing the lifting of heavy or light boxes. Objects that were described as matching the content of the sentence elicited larger hand and arm delay (i.e., the time after the object is grasped) relative to those described as mismatching, thus suggesting that weight information was activated.

While an inferred change of state implied by object weight is not at the focus of Altmann's and Ekves' (2019) IOH work, such kind of inference is entirely compatible with one of the central tenants of their theoretical account, which is that semantic memory (world knowledge) can inform constructed events. More specifically, in the IOH account, events are comprised of contingent object histories (i.e. a current object state constrains all other possible states of the object), whereby an activation of one object state reactivates its entire history during the comprehension of the language. Importantly, spatiotemporal contingencies between events also lead to the emergence of higher-order contingencies across events, such as schemas (or events typical for a given situation) and scripts (or sequences of typical events for a given situation) that overlap during event representation. Thus, according to the IOH account, understanding a sentence "You drop a bowling ball on a tomato" should activate generalized event knowledge about the objects mentioned in the sentence and schema knowledge about likely chains of events in the context of dropping (given what we know about the weight and fragility of each object type). Therefore, the present research raises the following important question: is shape information updated differently when object state-change is implied by another object's weight rather than different locations or actions?

The present research

The present research used similar methods to those in a related study by Kang et al. (2019), but is different in several important respects. First, whereas in Kang et al. (2019) the shape of an object was implied by using two different verbs (e.g., dropped vs. chose), in our study the shape of an object was implied by virtue of a description of a heavy vs. light second object being dropped on it (e.g., a bowling ball vs. a balloon). Second, while in Kang et al. (2019) the object following a verb was easily "squashable" (e.g., ice cream), in our study it was not (e.g., sponge, dumbbell). Third, in Kang et al. (2019) sentences were in the third-person condition and in our study sentences were in the second-person "you" condition. Finally, and most importantly, whereas in Kang et al. (2019) the object dropped was the one participants had to verify, in our experiments participants had to verify a "squashable" object onto which another "unsquashable" heavy or light item was dropped (e.g., "You drop a bowling ball/a balloon on a tomato"). Thus, in our study verbs and, more generally, linguistic cues are not the primary "drivers" of the updating of state information. Rather, these are the unmentioned, non-visual features of the situation (e.g., the weight of a bowling bowl when it is dropped on a tomato) defined by different objects coming together in time and space. These considerations are theoretically important in respect of understanding the conditions when multiple object state representations must be *simultaneously* activated during event comprehension.

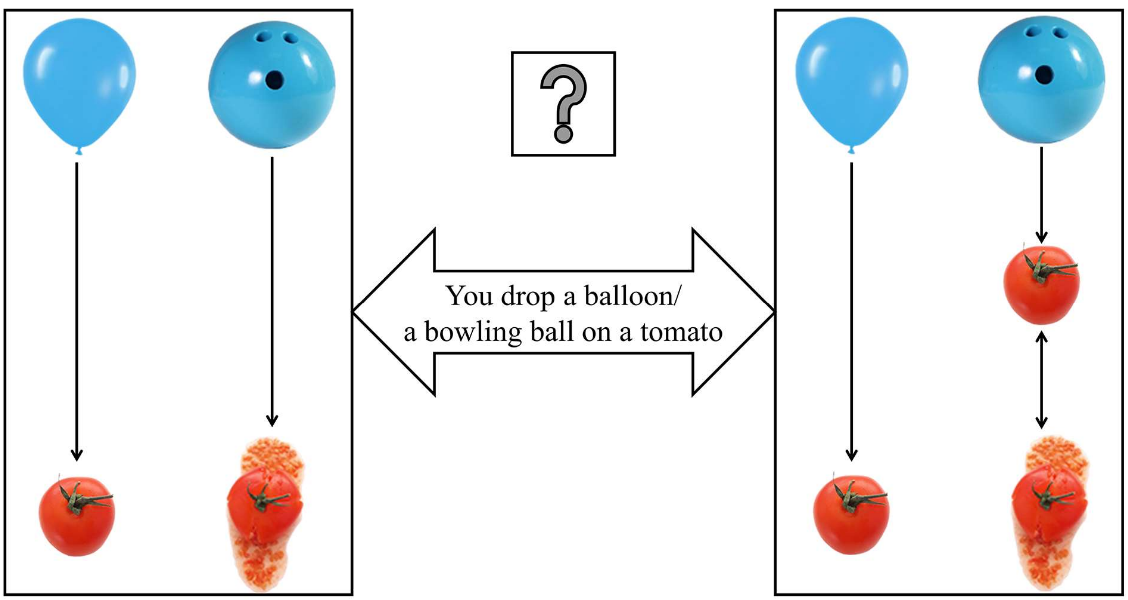

Figure 1. Two possible patterns of activation (on the left and on the right) that may facilitate the verification of pictured stimuli after reading sentences as the ones presented in the left-right arrow.

There are two hypotheses as to how the implied target object state may be represented after reading sentences such as “You drop a bowling ball on a tomato” and “You drop a balloon on a tomato”, which are summarized in Figure 1. The first possibility (arbitrarily referred to here as a “constant scenario”) is that only the consequences of the described action will be encoded, suggesting that at the time of reading how a heavy or a light item collides with a target object, participants form an immediate mental image of a target object in its context-specific form (a deformed or a non-deformed state, respectively). On this view, verification time should be shorter whenever the pictured target object matches the end state implied by the sentence. This prediction is based on the results of previous research when shape information was manipulated as a function of a location (e.g., see Zwaan & Pecher, 2012, for more information).

The second possibility (arbitrarily referred to here as a “competing scenario”) is that both the initial canonical and the end non-canonical states of the object would be equally integrated into the mental model. On this view, pictures depicting undeformed objects should not be verified faster after reading a sentence such as “You drop a balloon on a tomato” than a sentence such as “You drop a bowling ball on a tomato”. This is the case because comprehension of the dropping event in “You drop a bowling ball on a tomato” requires activation at “the tomato” of both the canonical state of a tomato and of the non-canonical deformed state - the consequence of the bowling ball dropping on it. On the contrary, pictures depicting deformed objects should be verified faster after reading a sentence such as “You drop a bowling ball on a tomato” than a sentence such as “You drop a balloon on a tomato”, precisely because the deformed state of an object is only implied by the “bowling ball” sentence. Such a prediction is in line with the results of an fMRI study of Hindy et al. (2012), where a competition effect was observed in sentences like “You stamp on a penny” vs. “You step on an egg”, as well as a new theory of events as intersecting object histories (Altmann &

Ekves, 2019) that attributes importance to both the initial and the end object states during event representation.

Experiment 1

Participants read sentences such as "You drop a bowling ball on a tomato" or "You drop a balloon on a tomato" and then decided whether the subsequently pictured object was mentioned in the sentence (see Figure 2, for samples of picture stimuli used). This experiment was designed to determine (1) whether weight information is considered when representing a target object shape and (2) whether sentence processing affects picture verification in line with a "constant" scenario or a "competing" scenario outlined earlier.

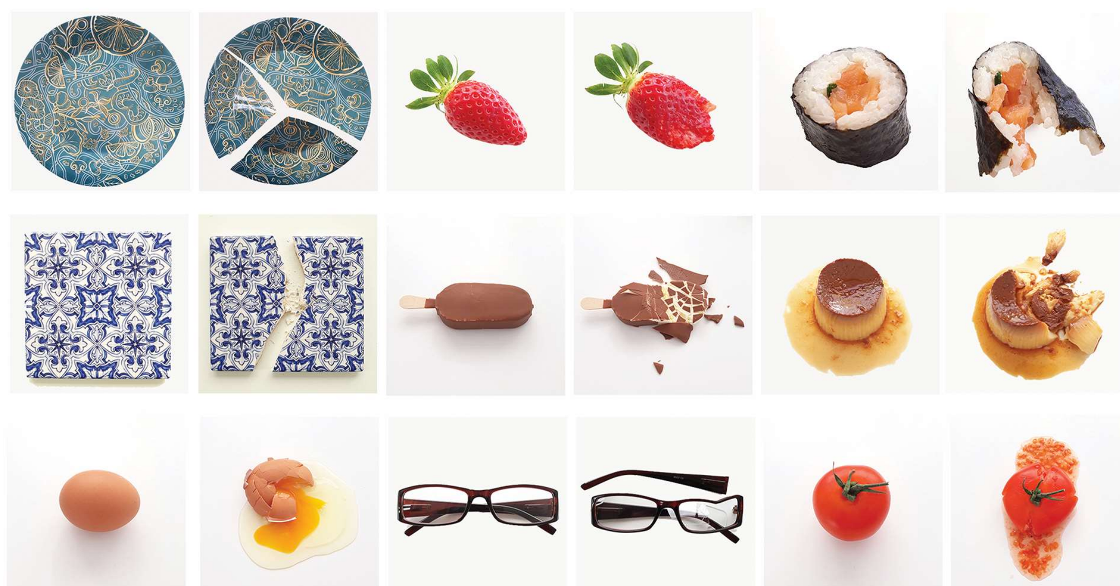

Figure 2. Sample picture stimuli used in Experiments 1-7.

Method

Sample size and ethical requirements

Power analysis was conducted in G*Power. Running a power analysis on a repeated measures ANOVA, a power of 0.90, an alpha level of 0.05, and a medium (η_p^2 = .06) effect size (Faul, Erdfelder, Lang, and Buchner, 2007), we expected to need at least 77 participants for each experiment. An estimate of medium effect size for power analysis is based on the results of Hoeben et al. (2019) and Sato et al. (2013) whose reported effect sizes of major effects were medium and large, respectively. To account for low accuracy scores and compliance with the task requirements, we always attempted to have at least 90 participants in each of our experiments. In line with the ethical guidelines of the host institution, participants from all seven experiments gave informed consent prior to participation and were fully debriefed about the purpose of the study upon completion.

Participants

One hundred and four native Portuguese-speaking university students took part in Experiment 1 in exchange for course credit. The responses of five participants were discarded for having accuracy <80% on the main task (four participants) or answering <50% of the comprehension questions correctly (one participant). Overall, the results of Experiment 1 are based on data from 99 participants (M_{age} = 23.33, SD_{age} = 4.53), of whom 79 were females.

Materials

Twenty-four experimental sentence pairs were created describing an action that involved dropping either a bowling ball or a balloon on objects that are unlikely to withstand a great deal of applied force without deformation (e.g., strawberry, light bulb). Thus, participants processed sentences involving differently weighted objects that implied contrasting degrees of applied force (i.e., bowling ball and balloon). The reason for choosing only a bowling ball and a balloon as the objects being dropped was to maximize control over other visual features that were shown to influence participants' expectations about object weight (e.g., size-weight illusion: Brenner & Smeets, 1996; shape: Glover, 2004). All of the experimental sentences were followed by a pictured object (e.g., an intact tomato or a squashed tomato) mentioned in the sentence and required "yes" responses.

Nonetheless, in order to prevent participants from paying attention to the words "bowling ball" and "balloon", we constructed twice as many filler sentences. Twelve of these sentences were of the same format as experimental sentences, but involved the dropping of multiple objects: 10 sentences were followed by a pictured object not mentioned in the sentence and required “no” responses; and two sentences were followed by a pictured object mentioned in the sentence and required “yes” responses. Furthermore, 36 sentences were constructed that focused on the act of seeing rather than action (e.g., “You see how a puppy is playing with a ball”). Ten of these “visual” sentences were followed by a pictured object (e.g., an intact object or a deformed object) mentioned in the sentence and required “yes” responses; and 26 of “visual” sentences were followed by a pictured object (e.g., an intact object or a deformed object) not mentioned in the sentence and required “no” responses. Finally, 24 comprehension questions[1] were created to alert participants of the need to pay attention to the meaning of the sentences (e.g., “You dropped a fork on a plate?”). These questions, which were not primary dependent variables to us, appeared after half of filler items and required an even distribution of “yes” responses and “no” responses. Each participant saw 24 experimental sentence–picture pairs requiring “yes” responses, 12 filler pairs requiring “yes” responses, and 36 filler pairs requiring “no” responses. Thus, there were 36 sentence–picture pairs requiring “yes” and 36 requiring “no” responses.

Seventy-two same-sized (385x385 pixels) images were created to accompany the sentences. Twenty-four pictures were experimental pairs. Both members of each pair depicted the same object except for the version of the object used: undeformed (canonical) or deformed (non-canonical). The other 48 pictures were fillers, with half of the pictures depicting an undeformed version of an object and the other half depicting a deformed version of an object. Almost all experimental pictures were created for this experiment by taking pictures of real objects. Most of the filler pictures were found on the Internet.

Design and procedure

There were four lists of stimuli, with each experimental sentence-picture pair appearing in only one of the following conditions per list: heavy-non-canonical; heavy-canonical; light-non-canonical; and light-canonical. There were 6 trials for each condition. Each participant saw one list only and was randomly assigned to it. The idea of list was to counterbalance items and conditions, so that the same items that appeared in one sentence-picture condition for some participants were in the different sentence-picture condition for other participants. A 3-way interaction between list, picture type, and sentence type was not significant ($t < 2$ in estimates of fixed effects using linear mixed-effects modelling). Thus, list was not included as a factor in the reporting of statistical analyses due to its little theoretical relevance (Pollatsek & Well, 1995). This led to a 2 (sentence: heavy vs. light) × 2 (picture: canonical vs. non-canonical) within-participants design.

E-Prime 2.0 was used for stimulus presentation. The experiment began with six practice trials to ensure that participants understood the instructions. After each practice trial (but not main trials) participants saw different feedback screens based on whether correct or incorrect response was provided. Instructions warned participants that throughout the experiment they would be asked to respond to some comprehension questions, and hence need to read sentences attentively. Following previous similar research (e.g., Kang et al., 2019), each trial of the main part of the experiment started with a fixation cross in the middle of a computer screen for 1000 milliseconds. Then a sentence appeared at the center of the screen until participants pressed the Spacebar, thus indicating that they read and understood the sentence. After a spacebar press, the sentence was replaced by a fixation cross for 500 milliseconds, immediately followed by a picture of an object (in either a non-canonical or a canonical state) that was either mentioned or not in the preceding sentence. Participants indicated their decision by pressing an “S” button for a “yes” response and an “N” button for a “no” response.

Data treatment

Prior to analysis, and in all seven experiments, incorrect responses, filler items, and the data of participants with an overall accuracy <80% on the main task (i.e., participants were at least 80% accurate in indicating that a target object was mentioned in the sentence regardless of implied object state) and <50% on the comprehension questions were excluded. Second, response times (RTs) were checked for normality using Q-Q plots and histograms with normal curve. In all seven experiments RTs were positively skewed, and thus log10 transformation was applied to get normal distributions (e.g., Baayen, 2008). Finally, responses exceeding ±3 median absolute deviations (MAD) from the condition's median were removed. To calculate MAD, the formula $MAD = median\ (|xi - median(x)|)$ was used, where median (x) is the median of the distribution and MAD equals the median of the differences between individual observations xi and the distribution. ±3 MAD is considered to be a robust method of outlier treatment that is not affected by extremely high or extremely low values, and thus eliminates the need to set upper and lower cutoff points (see Leys, Ley, Klein, Bernard, & Licata, 2013, for more information). Most of the experiments from the sentence–picture verification task used the median for the analyses (see Pecher & Zwaan, 2012, for a discussion), and hence choosing the method of outlier treatment based on median absolute deviation seemed to us as the most optimal.

Data analysis

All statistical analyses were performed within the R programming environment version 4.0.0 (R Core Team, 2020) and several R packages. We used the "tidyverse" package (Wickham et al. 2019) for data processing; the "lme4" package (Bates, Mäechler, Bolker, & Walker, 2015) and "lmerTest" package (Kuznetsova, Brockhoff & Christensen, 2017) for main statistical analyses of accuracy and response times; the "report" (Makowski & Lüdecke, 2019) and "sjPlot" (Lüdecke, 2020) packages for reporting statistical results. R Markdown files were used to generate code and the analyses were "knit" into html files that contain our comments, code, and output. We used the default R "treatment" (or dummy coding) coding scheme, where each level of the categorical variable is contrasted to a specified reference level. In the present research, the "heavy" sentence condition and the "non-canonical" picture condition were set as reference categories. Given that the interpretation of lower-order effects (such as main effects) is affected by the presence of an interaction when fitting models using treatment contrasts (Singmann & Kellen, 2020), throughout the paper we reported the full model followed by two models aimed at extracting simple effects - one for the "canonical" picture condition and one for the "non-canonical" picture condition. If the presence of an interaction was not established, we removed the non-significant interaction term from the model and reran the analysis with two fixed effects only (i.e., sentence, picture).

Accuracy

Logistic mixed-effects regression with crossed random effects of participants and items was used to analyze accuracy scores. For both accuracy and response times analyses, we fitted the full variance-covariance structure of random effects (the so-called "maximal" model; Barr et al., 2013). The "maximal" model for the present research is the one with sentence, picture, and their interaction considered as fixed effects; random intercepts for participants and items; by-participants random slopes for sentence, picture, as well as the interaction term; a maximum likelihood estimation parameter; and an unstructured covariance matrix. Note, however, that no random slopes were specified for items as each participant gave only one response per individual test item (see Barr et al., 2013, for more information). If the "maximal" model failed to converge, we first checked whether the model converges with a random effects structure for which no slope-intercept correlation term is specified (to minimize risks of model reduction). Only when this did not help, we reduced the model by removing a random slope that makes a model fail to converge.

Response times

Linear mixed-effects models with crossed random effects of participants and items were used to analyze response times (Baayen et al., 2008). The advantage of using linear mixed effects models over traditional separate by-participants (commonly denoted as F_1) and by-items (commonly denoted as F_2) repeated-measures ANOVAs is that this method of statistical analysis (1) handles the crossing of two random factors simultaneously (Baayen et al., 2008) and (2) takes into account all individual RTs rather than just mean or median RTs for each participant (Baayen & Milin, 2010). Similar to accuracy analyses, we fitted the "maximal" model to predict RTs and reduced the complexity of random-effects structure only if the model failed to converge (in order to prevent unknown risk of anticonservativity).

Results and discussion

Data trimming for RTs

The removal of responses falling outside ±3 MAD from the relevant condition's median led to the loss of 6.13 % of observations.

Accuracy data

Participants' response accuracy was 97.6%. The "maximal" logistic mixed-effects regression model (estimated using ML and BOBYQA optimizer) to predict Accuracy converged successfully. The results showed that there was no interaction between sentences and pictures (*beta* = 2.28, *SE* = 3.71, $z = 0.61$, $p = .539$). Thus, we removed the non-significant interaction term from the model and reran the analysis with two fixed effects (sentence, picture) only. The results demonstrated no significant main effect of sentence type (*beta* = -0.35, *SE* = 0.44, $z = -0.79$, $p = .427$) and no significant main effect of picture type (*beta* = -0.67, *SE* = 0.55, $z = 1.20$, $p = .229$).

RT data

The data of major interest are provided in Figure 3. Following Baayen and Milin (2010), we present data using non-transformed means for the convenience to visualize effects in the millisecond scale. The "maximal" linear mixed-effects model (estimated using ML and BOBYQA optimizer) to predict RTs converged successfully. Most critical to our predictions was a significant interaction between sentences (heavy vs. light) and pictures (non-canonical vs. canonical), *beta* = -0.07, *SE* = 0.01, *t* = - 4.70, *p* < .001, 95% CI [-0.09, -0.04].

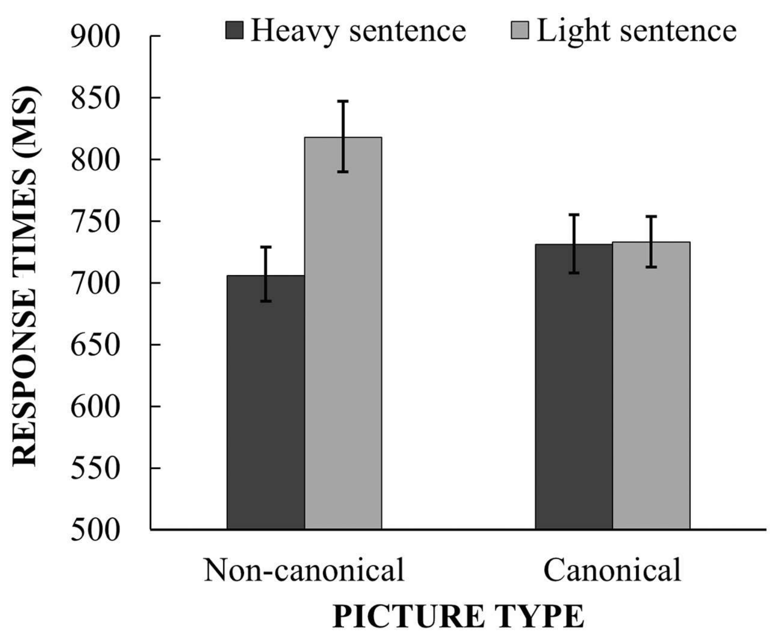

Figure 3. Mean non-transformed response times (in milliseconds) and error bars for verification of pictures depicting objects in either a non-canonical or a canonical state in Experiment 1. Error bars indicate 95% confidence intervals of the difference between the means of "heavy" and "light" sentences in each picture condition.

To investigate this interaction further, the data file was split by pictures and separate multilevel models on the "non-canonical" pictures and "canonical" pictures were conducted (Field, 2013). The models specified included a fixed effect of sentence type, a by-participant random slope for sentence type, and a random intercept for participants and items. The data showed that participants verified "non-canonical" pictures more quickly when preceded by a "heavy" sentence than when preceded by a "light" sentence, *beta* = 0.06, *SE* = 0.01, *t* = 6.86,

$p < .001$, 95% CI [0.04, 0.08]. However, participants did not verify “canonical” pictures faster when preceded by a “light” sentence than when preceded by a “heavy” sentence, *beta* = 0.00, *SE* = 0.01, *t* = -0.36, *p* = .721, 95% CI [-0.02, 0.01].

Overall, these results are consistent with the “competing” scenario outlined earlier, which suggests that the initial target object representation cannot be completely overwritten when the context requires activating a different representation of object state. Thus, representations of an object’s initial and final states were simultaneously active in the “heavy” (“bowling ball” sentence) condition.

Experiment 2

Experiment 2 was designed to replicate the results of Experiment 1 with more sentence and picture stimuli, as well as to lend more credence to our argument that processing of an object in its crushed state relies on the re-activation of an object’s history – the trajectory of the past intact state that led to the current one. To this end, in Experiment 2 we added a “non-action” condition where participants processed simple sentences in which the verb solely denoted visual perception of the target object (e.g., “You see a tomato”) followed by either a “non-canonical” or a “canonical” pictured version of the object mentioned in the sentence. The purpose of this condition was to unravel what happens when both object states of a target object (e.g., squashed tomato vs. intact tomato) do not contradict sentence content. In line with the results observed in Experiment 1 of Kang et al. (2019), our prediction was that response latencies should be faster for “canonical” pictures than for “non-canonical” pictures after reading a non-action sentence like “You see a tomato”, precisely because prototypical object state should have an advantage in response times compared to the modified state.

Method

Participants

One hundred and eight native Portuguese-speaking university students participated in the experiment in exchange for course credit. The responses of seven participants were excluded for having accuracy <80% on the main task (four participants) or answering <50% of the comprehension questions correctly (three participants). Hence, the results of Experiment 2 are based on data from 101 participants (M_{age} = 22.42, SD_{age} = 4.23), of whom 79 were females.

Materials

The critical sentences and pictures were the same as in Experiment 1, except that an additional 12 sentences were constructed for the "non-action" condition and 12 new picture pairs were added. Each participant saw 36 experimental sentence–picture pairs requiring "yes" responses, 12 filler pairs requiring "yes" responses, and 36 filler pairs requiring "no" responses. Thus, there were 48 sentence–picture pairs requiring "yes" and 36 requiring "no" responses.

Design and procedure

To have a counterbalanced design, six lists were created and each list included one of six possible versions (3 sentences: heavy, light, non-action; 2 pictures: canonical, non-canonical) for each object. There were 6 trials for each experimental condition. The procedure was the same as in Experiment 1.

Results and discussion

Data trimming for RTs

The removal of responses falling outside ±3 MAD from the relevant condition's median led to the loss of 5.28 % of observations.

Accuracy data

Participants' response accuracy was 97.9%. Given that in Experiment 2 one of the variables (i.e., sentence) had more than 2 levels, we performed a likelihood ratio test that compares the likelihood of one model to the likelihood of another model in order to determine whether the 3x2 interaction was significant. A likelihood ratio test of the "maximal" model with fixed effects of sentence, picture, and their interaction against the "simplified" model with fixed effects of sentence and picture revealed no significant difference between models (χ^2 (13) = 10.32, p = .667), thus suggesting that there was no evidence for the 3x2 interaction between sentence type ("heavy", "light", "non-action") and picture type ("canonical", "non-canonical"). Thus, we used the "simplified" model with no interaction term in the reporting of statistical analyses. The results of this model revealed that simple effects of "light" sentences (*beta* = 0.29, *SE* = 0.45, z = 0.65, p = .518) and non-action" sentences (*beta* = 0.60, *SE* = 0.45, z = 1.34, p = .179) were not significant relative to the referent level (i.e., heavy sentences). Furthermore, by making a non-action sentence condition as a referent category, we found that "non-action" sentences were not verified more accurately than "light" sentences (*beta* = -0.31, *SE* = 0.53, z = -0.59, p = .554). Finally, there was no significant main effect of picture type (*beta* = -0.45, *SE* = 0.46, z = -0.98, p = .329).

RT data

The results of major interest are presented in Figure 4. A likelihood ratio test of the "maximal" model with fixed effects of sentence, picture, and their interaction against the best converging model[2] with fixed effects of sentence and picture revealed a significant difference between the models (χ^2 (12) = 81.11, $p < .001$), thus suggesting that there was a strong evidence for the 3x2 interaction between sentence type ("heavy", "light", "non-action") and picture type ("canonical", "non-canonical").

Consistent with our reasoning, follow-up analyses showed that verification times for "canonical" pictures were significantly faster than verification times for "non-canonical" pictures after reading "non-action" sentences, *beta* = -0.05, *SE* = 0.01, t = -6.08, $p < .001$, 95% CI [-0.06, -0.03]. Furthermore, in line with the results of Experiment 1, the segregation of the items by pictures showed that "non-canonical" pictures were responded to more quickly when preceded by a "heavy" sentence than when preceded by a "light" sentence, *beta* = 0.04, *SE* = 0.01, *t* = 5.38, *p* < .001, 95% CI [0.03, 0.06]; but "canonical" pictures were not responded to significantly faster when preceded by a "light" sentence than when preceded by a "heavy" sentence, *beta* = -0.01, *SE* = 0.01, *t* = -0.79, *p* = .429, 95% CI [-0.02, 0.01].

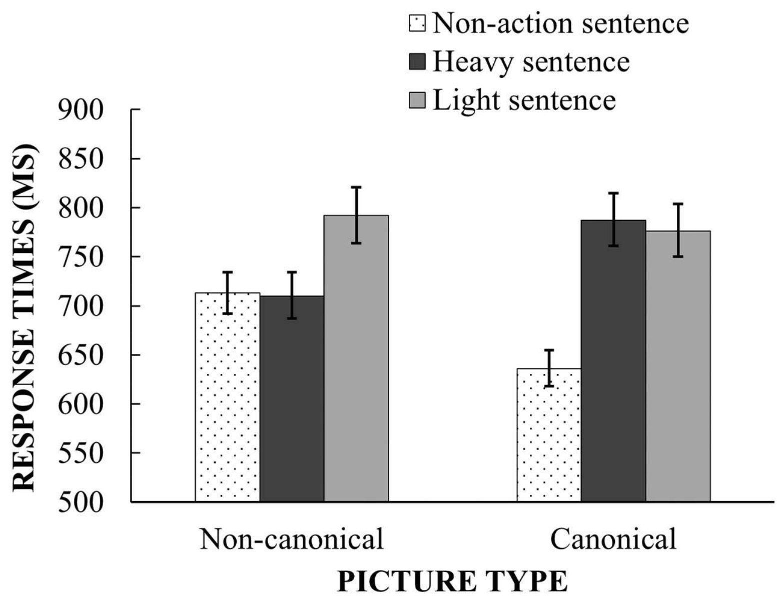

Figure 4. Mean non-transformed response times (in milliseconds) and error bars for verification of pictures depicting objects in either a non-canonical or a canonical state in Experiment 2. Error bars indicate 95% confidence intervals of the difference among the means of "heavy", "light", and "non-action" sentences in each picture condition.

There were also some other interesting effects in the analyses of response times. More specifically, by setting a "non-action" sentence condition as a referent category we found that verification times for pictures depicting "canonical" objects were much faster when preceded by a non-action sentence than when preceded by a "light" sentence, *beta* = 0.08, *SE* = 0.01, *t* = 8.71, *p* < .001, 95% CI [0.06, 0.10]. This effect is not surprising considering the varying degree of task demands for these sentence conditions: non-action sentences mentioned one object always occurring in the subsequent picture (e.g., "You see a *tomato*") and "light" sentences mentioned two objects equally likely to occur in the subsequent picture (e.g., “You drop a *balloon* on a *tomato*”). In addition, such a result is consistent with prior research showing that the construction of mental simulations is delayed when sentences are more complex (Kaup et al., 2006). However, verification times for pictures depicting “non-canonical” objects were nearly identical when preceded by both a “heavy” sentence and a non-action sentence, *beta* = 0.00, *SE* = 0.00, *t* = 0.03, *p* =. 975, 95% CI [-0.02, 0.02]. Presumably there was a response facilitation arising from the processing of “heavy” sentences in the “non-canonical” picture condition. If this were not the case, then the results for easier “non-action” sentences in the “non-canonical” picture condition should have been very much similar to the results for non-action sentences in the “canonical” picture condition. Thus, when looking at congruency effects for canonical and non-canonical picture conditions separately, the results replicate those from Experiment 1.

Nonetheless, the pattern of results as a whole is not fully consistent with the findings from Experiment 1. For some reason response times for canonical objects after “heavy” and “light” sentences in Experiment 2 were on average longer than those observed in Experiment 1 (see Figures 3 and 4). That is, if there was a competition of object states in the process of language comprehension, then no significant match advantage for the original object state (i.e., intact tomato) should have been observed relative to the modified object state (i.e., squashed tomato) in the substantial change (“heavy” sentence) condition. We conducted Experiment 3 to further address this issue.

Experiment 3

Experiment 1 showed that pictures depicting “non-canonical” objects were verified faster after participants read a sentence describing the action of dropping a bowling ball on a target object than a sentence describing the action of dropping a balloon. Experiment 2 replicated the above finding and provided further support for our claim that prototypical object information, which is initially activated (e.g., an intact tomato), is not completely overwritten or inhibited when the context requires the activation of a different representation (e.g., of a deformed tomato) and therefore it can still affect picture verification. However, response times for canonical objects after “heavy” and “light” sentences in Experiment 2 were on average longer than those observed in Experiment 1 (see Figures 3 and 4). We reasoned that such variable results could be indicative of the presence of a “hidden moderator”. More specifically, we suspected that participants could have represented different kinds of balloons (e.g., air-filled vs. helium-filled) while reading the target sentences. For example, it could be that “balloon” sentences led some participants to mentally represent an upward direction of a described object’s motion (i.e., the case of a helium-filled balloon) rather than a downward direction of a described object’s motion (i.e., the case of an air-filled balloon), which could, in turn, have some consequences for the speed with which participants verified pictured targets. To test this possibility, we disambiguated the meaning of an item being dropped by now presenting participants with the sentences like “You drop a balloon *full of air* on a tomato”.

In addition, it is also possible that verification times of “canonical” pictures were different because the responses were made in the presence (when taken experiment as a whole) of non-action sentences inviting participants to visualize the described scene (e.g., “You see a tomato”). Therefore, we replaced “non-action” sentences from Experiment 2 (e.g., “You see a tomato”) with control sentences, which were identical to critical “heavy” and “light” sentence stimuli, except that the preposition “on” was replaced with the preposition “near” (e.g., “You drop a bowling ball/a balloon *near* a tomato”). The idea was to (1) check if the absence of “non-action” sentences would change the pattern of results and (2) to rule out the possibility that participants simply learned to associate a bowling ball with a “deformed” pictured stimulus and a balloon with an “undeformed” pictured stimulus while providing a response. Accordingly, we expected to show that (1) “non-canonical” pictures would be verified faster after a “heavy” sentence with a preposition “on” than a “heavy” sentence with a preposition “near”; and (2) “canonical” pictures would be verified equally fast after a “light” sentence with a preposition “on” and a “light” sentence with a preposition “near”. Thus, participants now processed the following experimental sentence types:

(1) You drop a bowling ball *on* a tomato;
(2) You drop a bowling ball *near* a tomato;
(3) You drop a balloon full of air *on* a tomato;
(4) You drop a balloon full of air *near* a tomato.

It is important to note that although we had four sentence types in total, a full-factorial design was used only for “heavy” and “light” sentences with a preposition “on” (as in previous two experiments). “Bowling ball” sentences with a preposition “near” were used only in conjunction with a picture condition depicting “non-canonical” objects; and “balloon” sentences with a preposition “near” were used only in conjunction with a picture condition depicting “canonical” objects. Crucially, however, the same items that appeared in the “on” condition for some participants were in the opposite “near” condition for other participants. Each participant saw 36 experimental sentence–picture pairs requiring “yes” responses, 12 filler pairs requiring “yes” responses, and 36 filler pairs requiring “no” responses. We did not use a full-factorial design for control sentences as such analyses would involve a comparison across mismatching trial types (e.g., verification of “non-canonical” pictures after “light” sentences with a preposition “on” and “light” sentences with a preposition “near”), which was of little theoretical interest to us. Thus, there were three sentence types for each picture condition (“light on”, “light near”, “heavy on” for “canonical” pictures; “heavy on”, “heavy near”, “light on” for non-canonical pictures). Given these limitations with the fixed-effect model matrix, simple effects were computed using all sentence types and the interaction term was computed using sentences with a preposition "on" only.

Method

Participants

One hundred and ten native Portuguese-speaking university students participated in the experiment in exchange for course credit. The responses of 10 participants were excluded for having accuracy <80% on the main task (seven participants) or having only one valid response in one of the experimental conditions (three participants). Thus, the results of Experiment 3 are based on data from 100 participants (M_{age} = 20.52, SD_{age} = 5.02), of whom 82 were females.

Materials

The critical sentences and pictures were the same as in Experiment 2, except for two changes. First, non-action sentences like "You see a tomato" were replaced by control sentences like "You drop a bowling ball *near* a tomato". Second, sentences involving the balloon were changed from "You drop a balloon on a tomato" to "You drop a balloon *full of air* on a tomato".

Design and procedure

Six lists were created for each object to counterbalance items and conditions. There were 6 trials for each condition. The procedure was the same as in previous two experiments.

Results and discussion

Data trimming for RTs

The removal of responses falling outside ±3 MAD from the relevant condition's median led to the loss of 5.05 % of observations.

Accuracy data

Participants' response accuracy was 97.3%. The "maximal" model did not converge successfully in Experiment 3. Therefore, the random effects structure was simplified and we fitted the "reduced" logistic mixed-effects regression model (estimated using ML and BOBYQA optimizer) to predict accuracy with sentence, picture, and their interaction as fixed effects; random intercepts for participants and items; and by-participants random slopes for sentence and picture (no interaction term). The results showed that there was no interaction between sentences and pictures (*beta* = 1.58, *SE* = 1.02, *z* = 1.55, *p* = .122). Thus, we removed the non-significant interaction term from the model and reran the analysis with two fixed effects (sentence, picture) only. The results showed that simple effects of "heavy near" sentences (*beta* = 0.24, *SE* = 1.18, *z* = 0.20, *p* = .841), "light" sentences (*beta* = 0.08, *SE* = 0.57, *z* = 0.14, *p* = .888), and "light near" (*beta* = 0.46, *SE* = 0.90, *z* = 0.52, *p* = .607) sentences were not significant relative to the referent level (i.e., heavy sentences). Furthermore, by making a "heavy near" sentence condition as a referent category, we found that "heavy near" sentences were not verified more accurately than "light" sentences (*beta* = -0.16, *SE* = 1.19, *z* = -0.13, *p* = .896) and "light near" sentences (*beta* = 0.23, *SE* = 1.41, *z* = 0.16, *p* = .872). By making a "light" sentence condition as a referent category, we found that "light" sentences were not verified more accurately than "light near" sentences (*beta* = 0.38, *SE* = 0.92, *z* = 0.42, *p* = .677). Finally, there was also no significant main effect of picture type (*beta* = -0.57, *SE* = 0.56, *z* = -1.02, *p* = .310).

RT data

The results of major interest are presented in Figure 5. Linear mixed-effects model analyses showed that a random slope for the sentence by picture interaction did not add to the model[3], and thus the results are based on the model (estimated using ML and BOBYQA optimizer) that included sentence, picture, and their interaction as fixed effects; random intercepts for participants and items; and by-participants random slopes for sentence and picture (no interaction term). Most central to our prediction, there was a significant interaction between sentence type (heavy vs. light) and picture type (canonical vs. non-canonical), *beta* = -0.04, *SE* = 0.01, *t* = -3.12, *p* =.002, 95% CI [-0.07, -0.02].

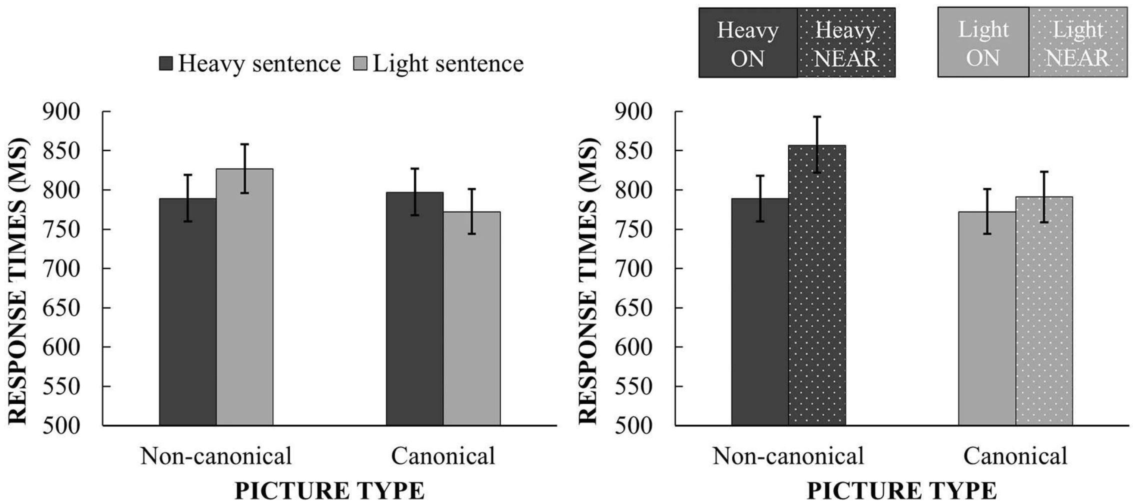

Figure 5. Mean non-transformed response times (in milliseconds) and error bars for verification of pictures depicting objects in either a non-canonical or a canonical state in Experiment 3. The left graph represents the difference in response times between the means of "heavy" and "light" sentences with preposition ON in each picture condition. The right graph represents the difference in response times between the means of (1) "heavy" sentences with preposition ON and "heavy" sentences with preposition NEAR in the non-canonical picture condition; and (2) "light" sentences with preposition ON and "light" sentences with preposition NEAR in the canonical picture condition.

We segregated the items by pictures to investigate this interaction further. As shown in Figure 5 (the left graph), "non-canonical" pictures were responded to significantly faster when preceded by a "heavy" sentence than when preceded by a "light" sentence (*beta* = 0.02, *SE* = 0.01, *t* = 2.54, *p* = .011, 95% CI [0.01, 0.04]); but "canonical" pictures were not responded to significantly faster when preceded by a "light" sentence than when preceded by a "heavy" sentence (*beta* = -0.02, *SE* = 0.01, *t* = -1.88, *p* = .060, 95% CI [-0.04, 0.00]).

Furthermore, as demonstrated in Figure 5 (the right graph), "non-canonical" pictures were responded to more quickly when preceded by a "heavy" sentence with a preposition "on" than when preceded by a "heavy" sentence with a preposition "near" (*beta* = 0.03, *SE* = 0.01, $t = 3.35$, $p = .001$, 95% CI [0.01, 0.05]); and "canonical" pictures were responded to almost equally fast when preceded by a "light" sentence with a preposition "on" and a "light" sentence with a preposition "near" ($b = 0.01$, *SE* = 0.01, $t = 1.19$, $p = .235$, 95% CI [-0.01, 0.03]).

Overall**,** the results replicate those of the previous two experiments with regards to the "non-canonical" picture condition and rule out the possibility that the observed congruency effects can be solely explained by lexical associations (i.e., the expectation that when I read a "bowling ball" something is necessarily going to get squashed). Finally, the data showed that after disambiguating the type of a balloon implied by the sentence (i.e., "balloon full of air" rather than just "balloon") and excluding non-action sentences the results were comparable with those from Experiment 1. Given that adding disambiguating information to the "balloon" condition did not change the overall pattern of results relative to Experiment 1, we are inclined to think that the inconsistency between Experiments 1 and 2 was due to the presence of non-action sentences. To lend further credence to this idea, we conducted further experiments in which a non-action condition was not included.

Experiment 4

Experiments 1 to 3 are clear in demonstrating that non-visual features of the situation (e.g., the weight of a bowling bowl when it is dropped on a tomato) are taken into account when representing object state (e.g., a squashed tomato). Furthermore, the data suggest that both the initial and the end states of an object are encoded, thus pointing to the strength of the initially activated representation (e.g., an intact tomato). Thus, the data are consistent with the results of Kang et al. (2019) who demonstrated that there is a competition of object states during event representation. Several questions remain, however. For example, could it be that verification latencies for “canonical” pictures in Experiments 1 to 3 were inhibited just because participants found it very hard to associate balloons with an action of dropping? Similarly, could it be that verification latencies for “non-canonical” pictures in Experiments 1 to 3 were facilitated just because participants found it very easy to associate bowling balls with an action of dropping? That such objects as bowling balls can be associated with an action of dropping is supported by research on object affordances showing that the action the object evokes may get activated independently of the described action (Tipper et al., 2006). To address these questions, in Experiment 4 we replaced “bowling ball” sentences with “brick” sentences (heavy condition) and “balloon full of air” sentences with “bath sponge” sentences (light condition). We thought that such objects as bricks and bath sponges cannot lead to any other representation of motion direction other than downward in the context of the action of dropping. Furthermore, we reasoned that action affordance effects are unlikely to be stronger for bricks than for bath sponges, given that the act of dropping does not closely resemble the situation of their natural use. If the effects observed in Experiments 1 and 3 are replicated, it would lend more credence to the claim that the encoding of both the initial and the end object states routinely occurs during sentence processing.

Method

Participants

One hundred and thirty native Portuguese-speaking university students participated in the experiment. The responses of 16 participants were excluded for having accuracy <80% on the main task (14 participants), answering less than 50% of comprehension questions correctly (1 participant), or having only one valid response in one of experimental conditions (1 participant). Thus, the results of Experiment 4 are based on data from 114 participants (M_{age} = 20.78, SD_{age} = 4.91), of whom 98 were females.

Materials

The critical sentences and pictures were the same as in Experiment 3, except for the following two changes. First, "bowling ball" sentences were replaced by "brick" sentences such as "You drop a brick on/near a tomato". Second, "balloon full of air" sentences were replaced by "bath sponge" sentences such as "You drop a bath sponge on/near a tomato".

Design and procedure

Design was the same as in Experiment 3. There were 6 trials for each condition. Procedure was the same as in previous three experiments.

Results and discussion

Data trimming for RTs

The removal of responses falling outside ±3 MAD from the relevant condition's median led to the loss of 3.98 % of observations.

Accuracy data

Participants' response accuracy was 96.7%. Similar to Experiment 3, the best converging logistic mixed-effects regression model (estimated using ML and BOBYQA optimizer) to predict accuracy was the one with sentence, picture, and their interaction as fixed effects; random intercepts for participants and items; and by-participants random slopes for sentence and picture (no interaction term). The results showed that the interaction between sentences and pictures was not significant (*beta* = 1.25, *SE* = 0.71, z = 1.76, p = .078). Thus, we removed the non-significant interaction term from the model and reran the analysis with two fixed effects (sentence, picture) only. The results showed that, relative to the referent level (i.e., heavy sentences), simple effects of "heavy near" sentences (*beta* = 0.22, *SE* = 0.61, z = 0.36, p = .718) and "light" sentences (*beta* = -0.03, *SE* = 0.39, z = -0.07, p = .947) were not significant, but the simple effect of "light near" (*beta* = 3.01, *SE* = 1.52, z = 1.98, p = .047) sentences was significant. Furthermore, by making a "heavy near" sentence condition as a referent category, we found that "heavy near" sentences were not processed more accurately than "light" sentences (*beta* = -0.25, *SE* = 0.61, *z* = -0.41, *p* = .685) and "light near" sentences (*beta* = 2.79, *SE* = 1.64, *z* = 1.70, *p* = .089). By making a "light" sentence condition as a referent category, we found that there was a statistical difference in how participants processed "light" sentences and "light near" sentences (*beta* = 3.09, *SE* = 1.55, *z* = 1.99, *p* = .047). Finally, there was no significant main effect of picture type (*beta* = 0.20, *SE* = 0.44, *z* = 0.46, *p* = .644).

RT data

The results of major interest are presented in Figure 6. The model (estimated using ML and BOBYQA optimizer) with the most maximal effects structure that converged (included sentence, picture, and their interaction as fixed effects; random intercepts for participants and items; and by-participants random slopes for sentence and picture) showed that there was a significant interaction between sentence type (heavy vs. light) and picture type (canonical vs. non-canonical), *beta* = -0.03, *SE* = 0.01, *t* = -2.36, *p* =.018, 95% CI [-0.05, 0.00].

As shown in Figure 6 (the left graph), the segregation of the items by pictures showed that "non-canonical" pictures were responded to significantly faster when preceded by a "heavy" sentence than when preceded by a "light" sentence, *beta* = 0.03, *SE* = 0.01, *t* = 3.11, *p* = .002, 95% CI [0.01, 0.04]. However, "canonical" pictures were not responded to significantly faster when preceded by a "light" sentence than when preceded by a "heavy" sentence, *beta* = 0.00, *SE* = 0.01, *t* = -0.02, *p* = .983, 95% CI [-0.02, 0.02]. Furthermore, and in line with the results from Experiment 3, Figure 6 demonstrates (the right graph) that "non-canonical" pictures were responded to more quickly when preceded by a "heavy" sentence with a preposition "on" than when preceded by a "heavy" sentence with a preposition "near", *beta* = 0.03, *SE* = 0.01, *t* = 3.60, *p* < .001, 95% CI [0.01, 0.05]; but "undeformed" pictures were not responded to significantly more quickly when preceded by a "light" sentence with a preposition "on" than when preceded by a "light" sentence with a preposition "near", *beta* = -0.01, *SE* = 0.01, t = -1.08, p = .282, 95% CI [-0.03, 0.01]. Thus, these data replicate the results of the previous experiments and demonstrate that the same pattern of responses is observed even when using differently-weighted items that do not evoke strong action-related affordance effects.

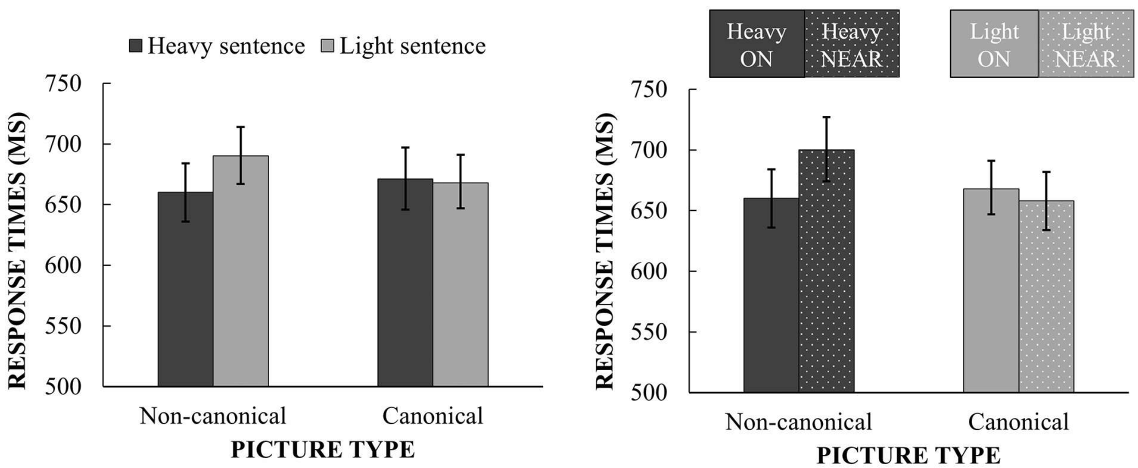

Figure 6. Mean non-transformed response times[4] *(in milliseconds) and error bars for verification of pictures depicting objects in either a non-canonical or a canonical state in Experiment 4. The left graph represents the difference in response times between the means of "heavy" and "light" sentences with preposition ON in each picture condition. The right graph represents the difference in response times between the means of (1) "heavy" sentences with preposition ON and "heavy" sentences with preposition NEAR in the non-canonical picture condition; and (2) "light" sentences with preposition ON and "light" sentences with preposition NEAR in the canonical picture condition.*

Experiment 5

The experiments presented so far indicate that both the initial and the end object states are integrated into the mental model during sentence processing. One interpretation of these results is that the canonical object representation, which is initially activated, can never be completely overwritten when shape is implied via the weight of an item that falls on a target object such as a tomato, precisely because the representation of a tomato's crushed state relies on the knowledge that the currently crushed tomato had existed in prior intact state. Alternatively, it is conceivable that the resultant state of the tomato is divorced from its initial state, but only when the verbal tense of the sentence indicates that the action has already happened (and is now over). Indeed, there remains a possibility that the canonical object representation could not be completely overwritten in Experiments 1 to 4 because the present tense of the sentence (e.g., "You drop a balloon/a bowling ball on tomato") implied that the deformation of a target object (e.g., a tomato) had yet to happen. Such a possibility is supported, in part, by previous eye tracking research showing that the tense and the aspect of the verb are used to determine the state of the object during the unfolding of the event (Altmann & Kamide, 2009; Knoeferle & Crocker, 2006). Furthermore, Kang et al. (2019) observed congruency effects for a canonical object state with past tense sentences (e.g., "The woman chose/dropped an ice-cream") but not with future tense sentences (e.g., "The woman will choose/drop an ice-cream") when shape information was manipulated by different verbs. Thus, we ran one more experiment to investigate whether the verbal tense of a sentence modulates the activation of the state of an object.

Method

Participants

Ninety native Portuguese-speaking university students participated in the experiment in exchange for course credit. The responses of four participants were excluded for having accuracy <80% on the main task. Additionally, the response of one participant had to be excluded for having unusually slow response times (>10 s). Hence, the results of Experiment 5 are based on data from 85 participants (M_{age} = 20.32, SD_{age} = 4.57), of whom 70 were females.

Materials

The critical pictures were the same as in Experiment 1. The sentences described the same objects as in Experiment 1, but the critical verb and the verbal tense of the sentence were changed. In Experiment 1 participants were presented with the present tense sentence like “You drop a bowling ball on a tomato” that was indeterminate with respect to whether the focus of the utterance was on the start state, an intermediary state, or the end state. In contrast, in Experiment 5 participants were presented with the subjectless, past tense sentence like “A bowling ball fell on a tomato” that shifted the focus of the utterance on the state of the world after the action had been completed. Finally, whereas before the object described as dropped was either a bowling ball and a balloon (Experiments 1, 2, and 3) or a brick and a sponge (Experiment 4), in Experiment 5 all of these objects were used to increase our confidence in the generalizability of the study. Each participant saw 24 experimental sentence–picture pairs requiring “yes” responses, 12 filler pairs requiring “yes” responses, and 36 filler pairs requiring “no” responses.

Design and procedure

The design was the same as in Experiment 1. There were 6 trials for each condition. The procedure was the same as in all previous experiments.

Results and discussion

Data trimming for RTs

The removal of responses falling outside ±3 MAD from the relevant condition’s median led to the loss of 2.61 % of observations.

Accuracy data

Participants’ response accuracy was 96.6%. The “maximal” logistic mixed-effects regression model (estimated using ML and BOBYQA optimizer) to predict Accuracy converged successfully. The results showed that there was no interaction between sentences and pictures (*beta* = 0.07, *SE* = 0.97, z = 0.07, p = .941). Thus, we removed the non-significant interaction term from the model and reran the analysis with two fixed effects (sentence, picture) only. The results demonstrated no significant main effect of sentence type (*beta* = 0.09, *SE* = 0.35, *z* = 0.26, *p* = .799) and no significant main effect of picture type (*beta* = 0.23, *SE* = 0.43, *z* = 0.55, *p* = .584).

RT data

The results of major interest are presented in Figure 7. The “maximal” model did not converge, and hence we fitted the “simplified” linear mixed-effects model (estimated using ML and BOBYQA optimizer) with uncorrelated intercept and slope for participants to predict RTs.

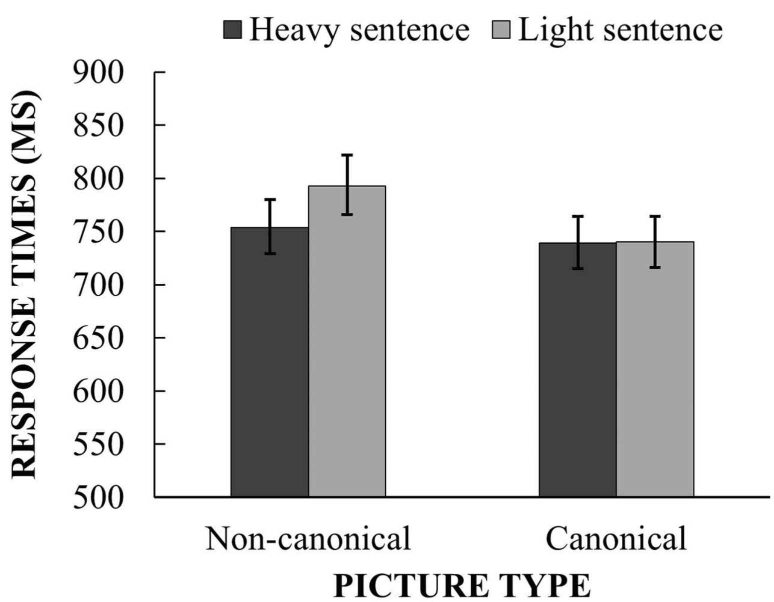

Figure 7. Mean non-transformed response times (in milliseconds) and error bars for verification of pictures depicting objects in either a non-canonical or a canonical state in Experiment 5. Error bars indicate 95% confidence intervals of the difference between the means of “heavy” and “light” sentences in each picture condition.

Most central to our prediction, there was a significant interaction between sentence type (heavy vs. light) and picture type (non-canonical vs. canonical), *beta* = -0.02*, SE* = 0.01, *t* = -2.03, *p* = .042, 95% CI [-0.05, 0.00].

The segregation of the items by pictures showed that “non-canonical” pictures were responded to more quickly when preceded by a “heavy” sentence than when preceded by a “light” sentence, *beta* = 0.03, *SE* = 0.01, $t = 3.25$, $p = .001$, 95% CI [0.01, 0.04]. However, “canonical” pictures were responded to equally fast when preceded by a “light” sentence and a “heavy” sentence, *beta* = 0.00, *SE* = 0.01, $t = 0.45$, $p = .653$, 95% CI [-0.01, 0.02]. These results consistently replicate the results of previous experiments (with an exception of Experiment 2, when responses were longer for canonical pictures), and thus provide compelling evidence that the object’s initial and final states are simultaneously active even after the verbal tense “forces” one to focus on the state of the world after the action had been completed.

Experiment 6

Experiment 5 replicated the results of the previous four experiments in that depictions of non-canonical objects after “heavy” sentences showed a match advantage, but pictures of canonical objects after “light” sentences did not. This suggests that the past tense of the sentence did not modulate the activation of object states. Notwithstanding the consistency of our results, two issues give us pause. First, it remains possible that the absence of a subject described in the involved event in Experiment 5 (“A brick/a bath sponge fell on a tomato”) could have affected the pattern of observed results. Situation models of language processing suggest that a subject (or protagonist) is one of the most critical components of the meaning in a sentence (Zwaan & Radvansky, 1998). Furthermore, there is direct empirical evidence showing that when the subject is omitted from the sentence, then image verification may be impaired (Sato & Bergen, 2013). Another question concerns the possibility that participants could have guessed over the course of an experiment that whenever a sentence described how the bowling ball/ brick and the balloon/sponge are being dropped on a target object, then there should be a “yes” response. Indeed, all of the experimental sentences in Experiments 1 to 5 were followed by a pictured object (e.g., an intact tomato or a squashed tomato) mentioned in the sentence and required "yes" responses, thus potentially leading to a reduced sensitivity to objects in the canonical state.

To address the aforementioned issues, we used an experimental design and materials almost identical to those in Experiments 1 to 5, except for the following two differences. First, we created 12 filler sentences in which the object dropped was the same as in the experimental sentences (i.e., a bowling ball, a brick, a balloon, a bath sponge), but the object subsequently shown mismatched sentence content, and thus required "no" responses. Second, we used past tense sentences all containing second person pronoun (e.g., "*You dropped* a brick/a bath sponge on a tomato"). Thus, apart from the verbal tense, sentences were substantially identical to those from Experiments 1 to 4.

If depictions of canonical objects show the effect in Experiment 6, then this would indicate that either a sentence frame or an experimental design explain the lacking effect of sentence type on responses to objects in the canonical shape (e.g., due to reduced sensitivity). If, however, the effects observed in Experiment 6 are similar to those from previous experiments, then this would indicate that "subjectless" sentences and previous experimental design had no effect on the observed pattern of results in Experiments 1 to 5.

Method

Participants

Due to a COVID-19 pandemic, 104 native Portuguese-speaking participants were recruited via Prolific Academic (Palan & Schitter, 2018) – an Internet platform aimed at connecting researchers with participants interested in taking part in research in exchange for monetary compensation of their time. To ensure that only Native Portuguese speakers were recruited, we entered the following custom prescreening criteria: Country of Birth = Portugal; Country of Residence = Portugal, and First (Native) Language = Portuguese. The responses of 13 participants were excluded for having accuracy <80% on the main task (5 participants) or answering <50% of comprehension questions correctly (8 participants). Thus, the results of Experiment 6 are based on data from 91 participants (M_{age} = 24.05, SD_{age} = 5.42), of whom 53 were males. With regards to occupation, 53 participants were students, 34 were workers, and four were unemployed. The experiment lasted approximately 10 minutes. Participants were compensated at a rate of £5.05 (British pounds) per hour.

Materials

The critical pictures were the same as in Experiment 5. The critical sentences described the same objects as in Experiment 5, except that in Experiment 6 we used past tense sentences all containing a second person pronoun (e.g., "*You dropped* a brick/a bath sponge on a tomato"). Finally, we replaced 12 filler sentences from Experiment 5 by 12 new sentences that described the same objects dropped as in experimental sentences, but which were followed by mismatching pictures, and thus required "no" responses. Overall, in Experiment 6 participants processed 44 sentences that described something dropped on a "squashable" object (24 sentences were experimental and 20 were fillers) and 28 sentences that focused on the act of seeing rather than action.

Design and procedure

The design was the same as in Experiment 5, except for the 12 new sentences mentioned above. Each participant saw 24 experimental sentence-picture pairs requiring "yes" responses, 12 filler sentence-picture pairs requiring "yes" responses, and 36 filler sentence-picture pairs requiring "no" responses. There were 6 trials for each condition. Thus, there were 36 pairs requiring "yes" responses and 36 pairs requiring "no" responses.

The procedure was substantially the same as in all previous experiments, except that the stimulus presentation was controlled by a web-based service PsyToolkit, which was designed for setting up, running, and analyzing reaction-time (RT) experiments and online questionnaires (Stoet, 2010, 2017). Recently, Kim et al. (2019) experimentally tested the reliability of this web-based service in comparison to a lab-based service E-Prime 3.0 in a complex psycholinguistic task. The researchers found that results obtained through Psytoolkit were in line with those obtained through E-Prime 3.0.

Results and discussion

Data trimming for RTs

The removal of responses falling outside ±3 MAD from the relevant condition's median led to the loss of 6.33 % of observations.

Accuracy data

Participants' response accuracy was 96.9%. The "maximal" logistic mixed-effects regression model (estimated using ML and BOBYQA optimizer) to predict Accuracy converged successfully. The results showed that the critical interaction between sentences and pictures was not significant (*beta* = 4.11, *SE* = 2.29, $z = 1.80$, $p = .072$). Thus, we removed the non-significant interaction term from the model and reran the analysis with two fixed effects (sentence, picture) only. The results demonstrated no significant main effect of sentence type (*beta* = -0.03, *SE* = 0.36, $z = -0.08$, $p = .941$). However, there was a significant main effect of picture type (*beta* = -1.32, *SE* = 0.54, $z = -2.44$, $p = .015$), which reflects the fact that participants were more accurate to verify "non-canonical" pictures than "canonical" pictures.

RT data

The results of major interest are presented in Figure 8. The "maximal" linear mixed-effects model (estimated using ML and BOBYQA optimizer) to predict RTs converged successfully. There was a significant interaction between sentence type (heavy vs. light) and picture type (non-canonical vs. canonical), *beta* = -0.02*, SE* = 0.01, *t* = -2.21, *p* = .027, 95% CI [-0.05, 0.00].

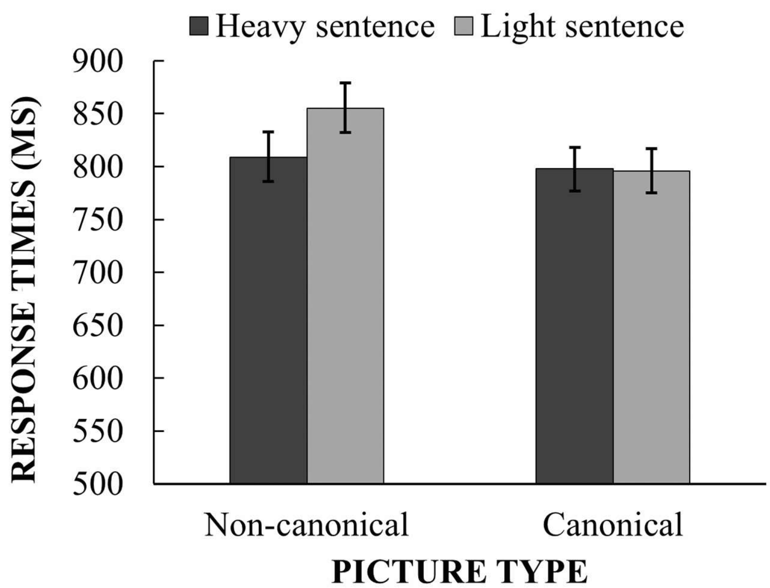

Figure 8. Mean non-transformed response times (in milliseconds) and error bars for verification of pictures depicting objects in either a non-canonical or a canonical state in Experiment 6. Error bars indicate 95% confidence intervals of the difference between the means of "heavy" and "light" sentences in each picture condition.

The segregation of the items by pictures showed that "non-canonical" pictures were responded to more quickly when preceded by a "heavy" sentence than when preceded by a "light" sentence, *beta* = 0.03, *SE* = 0.01, *t* = 3.08, *p* = .002, 95% CI [0.01, 0.04]. However, "canonical" pictures were responded to equally fast when preceded by a "light" sentence and a "heavy" sentence, *beta* = 0.00, *SE* = 0.01, *t* = 0.05, *p* = .957, 95% CI [-0.01, 0.01]. Thus, these results replicate those from the previous five experiments, suggesting that the resultant state of an object is not divorced from its initial state. Consequently, the "lacking" effect of sentence type on responses to objects in canonical shape was not due to a "subjectless" sentence or an experimental design.

Experiment 7

One potential criticism of Experiments 1 to 6 is that all of the critical sentences that describe something dropped on a “squashable” object included either a bowling ball and a balloon (Experiments 1, 2, and 3), a brick and a sponge (Experiment 4), or both pairs (Experiments 5 and 6). Given this situation, one may suspect the possibility of unusual processing on the part of the participants. While the repetitive use of the above stimuli allowed us to control for perceptual similarity of items (i.e., have the same size but differ in weight), it remains possible that such a repetition led to the partial overtness of the manipulation and, in turn, to the reduced sensitivity regarding objects in a canonical state. If comprehension of the dropping event requires activation of both the canonical intact state and the non-canonical deformed state of a target object in a real-life language comprehension scenario, then one should expect to see the same results with other heavy and lights objects regardless of their perceptual similarity. To ensure that the observed effects for canonical and non-canonical pictures are robust, we conducted a final experiment with multiple heavy-light pairs.

Method

Participants

We recruited 120 native Portuguese-speaking participants through Prolific Academic. Custom prescreening criteria were the same as in Experiment 6, except that in Experiment 7 we prevented participants from Experiment 6 from accessing the study. The responses of 15 participants were excluded for having accuracy <80% on the main task (6 participants), answering <50% of comprehension questions correctly (7 participants), or having only one valid response in one of the experimental conditions (2 participants). Hence, the results of Experiment 7 are based on data from 105 participants (M_{age} = 24.74, SD_{age} = 5.31), of whom 68 were males. With regards to occupation, 57 participants were students, 47 were workers, and 1 was unemployed. The experiment lasted approximately 10 minutes. Participants were compensated at a rate of £5.05 (British pounds) per hour.

Materials

In Experiment 7 we refined our study materials. First, all of the critical sentences that describe something dropped on a "squashable" target object now included only those items dropped that can be associated with the action of dropping in real-life contexts. Indeed, it might be argued that "heavy" sentences were overall responded to more quickly in previous experiments just because participants had a hard time drawing the causal link between the action of dropping and the object such as a "balloon" (e.g., balloons bounce or fly more frequently than drop). While we ruled out this possibility in Experiment 4 by replacing "balloon" with "bath sponge", we acknowledge that having only two pairs of items weakens the generalization scope. To address this concern, in Experiment 7 we thus used the following multiple heavy-light pairs: bowling ball/cotton ball; brick/sponge; dumbbell/cork; stone/ribbon; hammer/bank note; and frying pan/envelope (see Appendix A, for samples of critical sentences from Experiment 7). Importantly, similar to Experiment 6, we created 12 filler sentences in which all of the abovementioned dropped objects were followed by a pictured object not mentioned in the sentence and required "no" responses. Second, we excluded all nouns made up from more than one word used to describe a "squashable" object participants have to verify (e.g., plastic cup) to ensure that word complexity has no effect on the observed pattern of results. Pictures stimuli were taken from a set of pictures used in Experiments 2 to 4. Overall, in Experiment 7 participants processed 44 sentences that described something dropped on a "squashable" object (24 sentences were experimental and 20 were fillers) and 28 sentences that focused on the act of seeing rather than action.

Design and procedure

The design and procedure were the same as in Experiment 6.

Results and discussion

Data trimming for RTs

The removal of responses falling outside ±3 MAD from the relevant condition's median led to the loss of 5.05 % of observations.

Accuracy data

Participants' response accuracy was 95.6%. The "maximal" logistic mixed-effects regression model (estimated using ML and BOBYQA optimizer) to predict Accuracy converged successfully. Similar to all previous experiments, the results showed that the interaction between sentence type and picture type was not a significant predictor (*beta* = -0.72, *SE* = 0.94, z = -0.77, p = .442) for accuracy scores (see Table 1, for an overview of results in Experiments 1 to 7). Thus, we removed the non-significant interaction term from the model and reran the analysis with two fixed effects (sentence, picture) only. The results demonstrated no significant main effect of sentence type (*beta* = 0.02, *SE* = 0.32, z = 0.06, p = .956). However, there was a significant main effect of picture type (*beta* = -0.61, *SE* = 0.31, z = -1.98, p = .048), which reflects the fact that participants were more accurate to verify "non-canonical" pictures than "canonical" pictures.

RT data

The results of major interest are presented in Figure 9. The "maximal" linear mixed-effects model (estimated using ML and BOBYQA optimizer) to predict RTs converged successfully. Most central for our hypothesis, there was a significant interaction between sentence type (heavy vs. light) and picture type (non-canonical vs. canonical), *beta* = -0.04, *SE* = 0.01, t = -3.79, p < .001, 95% CI [-0.06, -0.02].

We segregated the items by pictures to investigate the interaction further. Similar to previous six experiments, we found that "non-canonical" pictures were responded to more quickly when preceded by a "heavy" sentence than when preceded by a "light" sentence, *beta*

= 0.03, *SE* = 0.01, *t* = 4.09, *p* < .001, 95% CI [0.02, 0.05]. However, “canonical” pictures were not responded to significantly faster when preceded by a “light” sentence than when preceded by a “heavy” sentence, *beta* = -0.01, *SE* = 0.01, *t* = -1.31, *p* = .190, 95% CI [-0.02, 0.00].

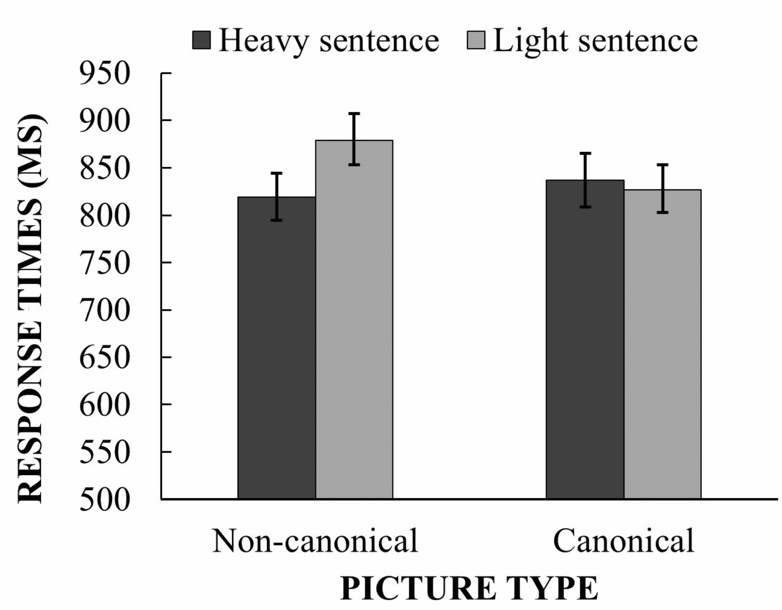

Figure 9. Mean non-transformed response times (in milliseconds) and error bars for verification of pictures depicting objects in either a non-canonical or a canonical state in Experiment 7. Error bars indicate 95% confidence intervals of the difference between the means of “heavy” and “light” sentences in each picture condition.

Thus, these results replicate those from the previous six experiments (see Table 1, for an overview of RT results in Experiments 1 to 7) and lend further credence to our claim that comprehension of the dropping event in a sentence like “You drop a bowling ball on a tomato” requires activation at “the tomato” of both the canonical state of a tomato and of the non-canonical deformed state - the consequence of the bowling ball dropping on it.

Table 1

Accuracy scores and mean response times (in milliseconds) for Experiments 1 to 7

	Picture Condition			
	Canonical		Non-canonical	
	Accuracy *M [95% CI]*	Response times *M [95% CI]*	Accuracy *M [95% CI]*	Response times *M [95% CI]*
		Experiment 1		
Heavy Sentence	0.97 [0.96, 0.98]	731 [708, 754]	0.98 [0.97, 0.99]	707 [685, 729]
Light Sentence	0.97 [0.96, 0.98]	733 [713, 754]	0.97 [0.96, 0.98]	818 [790, 847]
		Experiment 2		
Heavy Sentence	0.97 [0.96, 0.98]	787 [761, 815]	0.98 [0.97, 0.99]	710 [687, 734]
Light Sentence	0.98 [0.97, 0.99]	777 [750, 804]	0.97 [0.96, 0.98]	789 [762, 817]
Non-action	0.98 [0.97, 0.99]	636 [618, 656]	0.99 [0.98, 1.00]	713 [692, 735]
		Experiment 3		
Heavy Sentence	0.97 [0.96, 0.98]	798 [769, 828]	0.98 [0.97, 0.99]	790 [761, 820]
Light Sentence	0.96 [0.94, 0.98]	773 [745, 802]	0.97 [0.96, 0.98]	828 [797, 859]
Heavy Sent. (near)	-	-	0.97 [0.96, 0.98]	857 [823, 894]
Light Sent. (near)	0.98 [0.97, 0.99]	792 [760, 825]	-	-
		Experiment 4		
Heavy Sentence	0.96 [0.95, 0.97]	672 [647, 698]	0.97 [0.96, 0.98]	660 [637, 684]
Light Sentence	0.98 [0.97, 0.99]	669 [647, 691]	0.96 [0.94, 0.98]	691 [667, 715]
Heavy Sent. (near)	-	-	0.96 [0.95, 0.97]	701 [675, 727]
Light Sent. (near)	0.97 [0.96, 0.98]	659 [635, 683]	-	-
		Experiment 5		
Heavy Sentence	0.96 [0.94, 0.98]	735 [711, 760]	0.97 [0.95, 0.99]	754 [728, 780]
Light Sentence	0.97 [0.96, 0.98]	742 [717, 767]	0.97 [0.95, 0.99]	799 [771, 829]
		Experiment 6		
Heavy Sentence	0.96 [0.94, 0.98]	798 [777, 819]	0.98 [0.97, 0.99]	809 [786, 833]
Light Sentence	0.96 [0.94, 0.98]	796 [775, 817]	0.97 [0.96, 0.98]	855 [832, 879]
		Experiment 7		
Heavy Sentence	0.94 [0.92, 0.96]	836 [809, 865]	0.97 [0.96, 0.98]	819 [795, 844]
Light Sentence	0.95 [0.93, 0.97]	827 [803, 852]	0.97 [0.96, 0.98]	879 [853, 907]

General Discussion

The current research was conducted to address questions regarding the importance of object-state change for event representation during sentence processing. Central to these questions is empirical and theoretical evidence (Altmann & Ekves, 2019; Altmann & Mirković, 2009; Hindy et al., 2012; Solomon et al., 2015), according to which representations of different object states compete. In light of this evidence, we examined whether non-visual features of the situation (e.g., the weight of an item) are taken into account when representing target object shape. Furthermore, we investigated if the canonical object state continues to play an important role when the context requires the activation of a different object representation. The data support a conclusion that inferred changes of object state (e.g., the weight of a bowling bowl when it is dropped on a tomato) contribute to the updating of state information. Furthermore, the results show that understanding what happens to the tomato in a sentence like "You dropped a bowling ball on a tomato" requires an activation of the tomato in its final state and an activation of an earlier part of a tomato's trajectory. Such a conclusion follows from the results of all seven experiments that consistently demonstrated a match advantage for a "non-canonical" object state in the substantial change ("heavy" sentence) condition and no match advantage for a "canonical" object state in the no change ("light" sentence) condition, regardless of (1) whether the items dropped evoked/did not evoke action-related affordance effects, or (2) whether the tense of sentence implied/did not imply that the action is now over.

The results are particularly striking if one considers the effect of sentences without an action state (Experiment 2), which showed that in general "canonical" pictures are processed faster than non-canonical pictures but this difference is gone when canonical images are preceded by a "heavy" sentence. To make sense of this result, it is worth paying attention to the pattern of responses with regards to pictures in the original object state (i.e., intact tomato) and the modified object state (i.e., squashed tomato) in the substantial change ("heavy" sentence) condition. If our prediction regarding competition between object states is supported, then, similar to Kang et al. (2019), we should find no significant difference in participants' responses for the substantial change condition. And this is exactly what we observed (see Table 1) in almost all experiments. The results for this condition were only inconsistent to some extent in Experiment 2. At this point we are inclined to think that increased response times in the canonical picture condition in Experiment 2 were caused by "non-action" sentences inviting participants to visualize the described scene ("You see a tomato"). However, as we did not investigate the contribution of these sentences across experiments (i.e., the manipulation was within the same experiment), it is also possible that the results were different just because of random variability across experiments. A definite answer to this question must await further empirical investigation.

It is interesting to note that the analysis of response times in the same way as previous research (median RTs per condition; repeated-measures ANOVA) demonstrated a comparable pattern of results (see Appendix B): the differences that were significant using linear mixed-effects models were significant using repeated-measures ANOVA (all the time in the analysis by-participants F_1 and about half the time in the analysis by-items F_2). That is, the weight of the evidence across all seven experiments suggests that event comprehension requires the representation of both the intact and the modified states of the object – no matter what statistical method is being used to support this claim. Thus, our results are in line with Altmann's and Ekves' (2019) account of event representation, which posits that mental representations of an object's initial state are not deactivated but rather encoded into a situation model together with an object's end state.

While further work will be required to examine the extent to which the dynamics of the different object-state changes might unfold over the course of processing the sentence, the present study provides further constraints on theories of situation models (Johnson-Laird, 1983; Kintsch & van Dijk, 1978; Zwaan & Radvansky, 1998) in language comprehension. Our results indicate that each particular aspect of the episodic experience associated with an object differently defines how events should be integrated and updated into the situation model (see Zwaan et al., 1995, for a more in-depth discussion of how events can be indexed on such dimensions as causation, intentionality, protagonists, space, and time). Previous research showed that the final state of an object is more accessible when changes in location are implied (Mannaert et al., 2019; Sato et al., 2013; Zwaan et al., 2002). More recent research, however, demonstrated that when shape information is implied by using two different verbs (e.g., choose vs. drop), then mental representation of object state is dynamically updated, but in a more subtle way than could have been hypothesized. More specifically, the match advantage for a canonical object state was revealed only after processing past tense sentences (but not future tense sentences), thus suggesting that linguistic information modulates the activation of the relevant object representation (Kang et al., 2019). The present research suggests that when object weight is considered as a primary "driver" of the updating of state information, then the representations of an object's initial and resultant states are equally accessible. Therefore, the primary contribution of this study is that theories of cognition need to take account of those aspects of event meaning which are inferred from multiple objects coming together in space and time, rather than entailed by surrounding environment or lexical semantics.

There are a few critical factors that appear to determine the strength of activation of both initial and resultant object states during language comprehension. First, as discussed in the Introduction, in line with Altmann and Ekves (2019), we consider that an activation of a prior part of an object's trajectory depends on how useful or necessary it is to maintain that part of an object's trajectory. When event models are established around multiple objects coming together in time and space due to an external action, changes in the state of one object are casually related with changes in the state of another object. Consequently, to know what happened in the sentence "You dropped a bowling ball on an egg", one needs to encode the history of all the participating objects, which, among other things, includes both the initial and final states of an egg. In comparison, consider a study of Zwaan et al. (2002) in which the researchers instructed participants to read sentences about an egg in the fridge vs. in the skillet and found that verification times were shorter whenever the pictured object matched the final state implied by the sentence. In this study event models draw information from the surrounding context (i.e., location) in which an object is observed, and thus an object's trajectory is occluded. Therefore, even though an egg in its crushed state activates semantic knowledge about other possible states of an egg, an intact state of an egg is not a part of an object's trajectory we remain very sensitive to. Second, we assume that the competition of object states is most relevant for single events, and therefore our results are not comparable with studies that investigate multiple events (Mannaert et al., 2019). Third, we consider that one of the most critical factors in determining whether multiple object state representations are equally accessible during event comprehension is the critical verb used to describe action, precisely because object state-change is contingent on action. In our study we used the same verbs to describe action, and therefore participants should have had similar representations of the light/heavy objects' trajectories (i.e., downward movement) after reading both "You drop a brick on a tomato" and "You drop a sponge on a tomato". On the contrary, Kang et al. (2019) used two different verbs to describe object state-change, thus making it possible that some effects were driven by semantic associations between, for example, "drop/choose" actions and the "crushed/intact" perceptual properties of objects. Presumably it is this difference in sentence stimuli that explains why Kang et al. (2019) observed a match effect for objects in the canonical state in the minimal change ("choose" sentence) condition and we did not ("drop a balloon" sentence). Other than this difference, however, we consider our findings compatible with the Kang et al.'s (2019) work as they observed similar results to those reported in the present research. More specifically, they found that the initial and end states of objects were equally accessible in the future tense (e.g. "The squirrel will crack the acorn"), as well as that no match advantage is observed for the original object state (i.e., intact ice- cream) relative to the modified object state (i.e., squashed ice-cream) in the substantial change ("drop") condition. Importantly, our results are also in line with the results of an fMRI study of Hindy et al. (2012). In this research, although using different methods, researchers used the same verbs (e.g., "stamp on the penny / stamp on the egg") and concluded that there was a simultaneous activation of both objects states through observing a competition effect.

A limitation of the current study is that it does not allow us to make strong inferences as to the types of processes that underlie the activation of representational content. On the one hand, there is a wealth of behavioral and neuroimaging evidence that modality-specific systems are implicated in the representation of conceptual knowledge (Binder & Desai, 2011; Edmiston & Lupyan, 2017; Glenberg et al., 2008; Hauk et al., 2004; Horchak et al., 2014; Horchak et al., 2016; Ostarek & Huettig, 2017). On the other hand, there appears to be no direct experimental support for simulation-based accounts in the sentence-picture verification task at this point. Furthermore, a recent study of Ostarek et al. (2019) suggests that the findings from a sentence-picture verification task point to the rapid integration of implied visual information in sentence processing, but might be silent on the specific mental mechanisms underlying such integration. More specifically, the researchers tested a "shape simulation effect" with a visual noise manipulation (by using the materials from the study of Zwaan et al., 2002, where the shape of an object was implied via the location) and obtained no evidence that perceptual simulation underlies the match effect in the sentence-picture verification paradigm. With these caveats in mind, the current results should therefore be interpreted as providing evidence for the informational content that is activated in different conditions when object state information is implied via the weight of an item that falls on a target object. The functional role of the specific mental mechanisms underlying the rapid integration of implied visual, action, proprioceptive, and kinesthetic information during sentence processing has yet to be secured. This could be achieved by measuring processing at various stages (e.g., EEG method to assess temporal dynamic; Landau et al., 2010).

In conclusion, the present findings improve our insight into (1) how event information is updated into the situation model and (2) which representational content is encoded. Here, we have presented evidence that when changes of state are inferred (i.e., not driven by lexical semantics), both the initial and resultant states are equally accessible.

References

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

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*(1), 55–71. https://doi.org/10.1016/j.cognition.2008.12.005

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and Prediction in Human Sentence Processing. *Cognitive Science*, *33*(4), 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x

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. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. Harald, & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. https://doi.org/10.21500/20112084.807

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Bates , D. M., Mäechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Edmiston, P., & Lupyan, G. (2017). Visual interference disrupts visual knowledge. *Journal of Memory and Language*, *92*, 281–292. https://doi.org/10.1016/j.jml.2016.07.002

Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, *110*(4), 659–675. https://doi.org/10.1016/j.jecp.2011.06.009

Ferguson, H. J., Tresh, M., & Leblond, J. (2013). Examining mental simulations of uncertain events. *Psychonomic Bulletin & Review*, *20*(2), 391–399. https://doi.org/10.3758/s13423-012-0348-1

Field, A. (2013). *Discovering Statistics using IBM SPSS Statistics* (4th ed.). Sage Publications Ltd.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Lawrence Erlbaum Associates, Inc.

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., & Buccino, G. (2008). Processing abstract language modulates motor system activity. *Quarterly Journal of Experimental Psychology (2006)*, *61*(6), 905–919. https://doi.org/10.1080/17470210701625550

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, *41*(2), 301–307. https://doi.org/10.1016/S0896-6273(03)00838-9

Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object state-changes on event processing: Do objects compete with themselves? *The Journal of Neuroscience*, *32*(17), 5795–5803. https://doi.org/10.1523/JNEUROSCI.6294-11.2012

Hoeben Mannaert, L. N., Dijkstra, K., & Zwaan, R. A. (2019). How are mental simulations updated across sentences? *Memory & Cognition*, *47*(6), 1201–1214. https://doi.org/10.3758/s13421-019-00928-2

Horchak, O. V., & Garrido, M. V. (2020). Is Complex Visual Information Implicated During Language Comprehension? The Case of Cast Shadows. *Cognitive Science*, *44*(7), e12870.

Horchak, O.V., & Garrido, M. V. (2020, August 18). Data and Materials for "Dropping Bowling Balls on Tomatoes: Representations of Object State-Changes During Sentence Processing." Retrieved from osf.io/fd52u

Horchak, O. V., Giger, J.-C., Cabral, M., & Pochwatko, G. (2014). From demonstration to theory in embodied language comprehension: A review. *Cognitive Systems Research*, *29–30*, 66–85. https://doi.org/10.1016/j.cogsys.2013.09.002

Horchak, O. V., Giger, J. C., & Garrido, M. V. (2016). Action contribution to competence judgments: The use of the journey schema. *Frontiers in Psychology*, *7*, 448.

Horchak, O. V., Giger, J. C., & Pochwatko, G. (2014). Simulation of metaphorical actions and discourse comprehension. *Metaphor and Symbol*, *29*(1), 1-22.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press. https://hal.archives-ouvertes.fr/hal-00702919

Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., & Altmann, G. T. M. (2019). The influence of state change on object representations in language comprehension. *Memory & Cognition*. https://doi.org/10.3758/s13421-019-00977-7

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, *38*(7), 1033–1050. https://doi.org/10.1016/j.pragma.2005.09.012

Kim, J., Gabriel, U., & Gygax, P. (2019). Testing the effectiveness of the Internet-based instrument PsyToolkit: A comparison between web-based (PsyToolkit) and lab-based (E-Prime 3.0) measurements of response choice and response time in a complex psycholinguistic task. *PLoS ONE*, *14*(9). https://doi.org/10.1371/journal.pone.0221802

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

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481–529. https://doi.org/10.1207/s15516709cog0000_65

Kukona, A., Altmann, G. T. M., & Kamide, Y. (2014). Knowing what, where, and when: Event comprehension in language processing. *Cognition*, *133*(1), 25–31. https://doi.org/10.1016/j.cognition.2014.05.011

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1-26. doi: 10.18637/jss.v082.i13

Landau, A. N., Aziz-Zadeh, L., & Ivry, R. B. (2010). The Influence of Language on Perception: Listening to Sentences about Faces Affects the Perception of Faces. *The Journal of Neuroscience*, *30*(45), 15254–15261. https://doi.org/10.1523/JNEUROSCI.2046-10.2010

Lüdecke D (2020). sjPlot: Data Visualization for Statistics in Social Science. R package version 2.8.4, https://CRAN.R-project.org/package=sjPlot.

Makowski, D. & Lüdecke, D. (2019). The report package for R: Ensuring the use of best practices for results reporting. CRAN. Available from https://github.com/easystats/report.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Ostarek, M., & Huettig, F. (2017). A task-dependent causal role for low-level visual processes in spoken word comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(8), 1215–1224. https://doi.org/10.1037/xlm0000375

Ostarek, M., Joosen, D., Ishag, A., de Nijs, M., & Huettig, F. (2019). Are visual processes causally involved in "perceptual simulation" effects in the sentence-picture verification task? *Cognition*, *182*, 84–94. https://doi.org/10.1016/j.cognition.2018.08.017

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Short article: Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology*, *62*(6), 1108–1114. https://doi.org/10.1080/17470210802633255

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL 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*(5), 1224–1237. https://doi.org/10.1037/0278-7393.24.5.1224

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object shape and orientation do not routinely influence performance during language processing. *Psychological Science*, *24*(11), 2218–2225. https://doi.org/10.1177/0956797613490746

Sato, M., & Bergen, B. K. (2013). The case of the missing pronouns: Does mentally simulated perspective play a functional role in the comprehension of person? *Cognition*, *127*(3), 361–374. https://doi.org/10.1016/j.cognition.2013.02.004

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. https://doi.org/10.1515/langcog-2013-0022

Singmann, H., & Kellen, D. (2020). An Introduction to Mixed Models for Experimental Psychology. In D. H. Spieler & E. Schumacher (Eds.), *New Methods in Cognitive Psychology* (pp.4-31). New York, NY: Routledge.

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

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, *53*(1), 125–140. https://doi.org/10.1016/j.jml.2005.02.009

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Tipper, S. P., Paul, M. A., & Hayes, A. E. (2006). Vision-for-action: The effects of object property discrimination and action state on affordance compatibility effects. *Psychonomic Bulletin & Review*, *13*(3), 493–498. https://doi.org/10.3758/bf03193875

Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686, https://doi.org/10.21105/joss.01686

Zwaan, R. A. (1999). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science*, *8*(1), 15–18. https://doi.org/10.1111/1467-8721.00004

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. https://doi.org/10.1111/j.1467-9280.1995.tb00513.x

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 386–397. https://doi.org/10.1037/0278-7393.21.2.386

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PloS One*, *7*(12), e51382. https://doi.org/10.1371/journal.pone.0051382

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430

Footnotes

1. Participants' accuracy for comprehension questions was 82% in Experiment 1; 86% in Experiment 2; 91% in Experiment 3; 90% in Experiment 4; 88% in Experiment 5; 89% in Experiment 6, and 83% in Experiment 7. The cutoff point of only 50% accuracy on comprehension questions is explained by the fact that there wasn't enough evidence to conclude that participants' performance on the comprehension questions was always related to the accuracy on the main task. The results in all seven experiments are mixed. In Experiment 1, participants with an accuracy higher than 80% on the comprehension task had an accuracy of 98% on the main task; participants with an accuracy lower than 80% on the comprehension task had an accuracy of 97% on the main task, $\chi^2 = 3.16$, $p = .076$. In Experiment 2, participants with an accuracy higher than 80% on the comprehension task had an accuracy of 98% on the main task; participants with an accuracy lower than 80% on the comprehension task had an accuracy of 97% on the main task, $\chi^2 = 2.53$, $p = .112$. In Experiment 3, participants with an accuracy higher than 80% on the comprehension task had an accuracy of 97% on the main task; participants with an accuracy lower than 80% on the comprehension task had an accuracy of 96% on the main task, $\chi^2 = 1.11$, $p = .292$. In Experiment 4, participants with an accuracy higher than 80% on the comprehension task had an accuracy of 97% on the main task; participants with an accuracy lower than 80% on the comprehension task had an accuracy of 95% on the main task, $\chi^2 = 4.20$, $p = .040$. In Experiment 5, participants with an accuracy higher than 80% on the comprehension task had an accuracy of 97% on the main task; participants with an accuracy lower than 80% on the comprehension task had an accuracy of 96% on the main task, $\chi^2 = 0.93$, $p = .334$. In Experiment 6, participants with an accuracy higher than 80% on the comprehension task had an accuracy of 97% on the main task; participants with an accuracy lower than 80% on the comprehension task had an accuracy of 94% on the main

task, $\chi^2 = 8.50$, $p = .004$. In Experiment 7, participants with an accuracy higher than 80% on the comprehension task had an accuracy of 97% on the main task; participants with an accuracy lower than 80% on the comprehension task had an accuracy of 93% on the main task, $\chi^2 = 22.66$, $p <.001$.

2. The "maximal" model with no interaction term failed to converge, and thus we fitted the model with a random effects structure for which no slope-intercept correlation term is specified.
3. If random slopes that do not add to the model are not excluded, then the model fails to converge, thus attributing most of the variability to the participant's slope rather than the intercept (see Matuschek et al. 2017, for the discussion how to avoid the problem of overfitting the model to the data).
4. Picture verification times are globally shorter in Experiment 4 than in previous three experiments, perhaps because the participant sample for Experiment 4 consisted mostly of undergraduate psychology students who are used to taking part in reaction time experiments.

Appendix A

Samples of experimental sentences from Experiment 7

(sentences in original Portuguese language are provided in parentheses)

<u>"Heavy" sentences</u>

- You dropped a bowling ball on a tomato
  (*Deixaste cair uma bola de bowling num tomato*)
- You dropped a brick on a plate
  (*Deixaste cair um tijolo num prato*)
- You dropped a dumbbell on an iPhone
  (*Deixaste cair um halter num iPhone*)
- You dropped a stone on a blackberry
  (*Deixaste cair uma pedra numa amora*)
- You dropped a hammer on a papaya
  (*Deixaste cair um martelo numa papaia*)
- You dropped a frying pan on a bottle
  (*Deixaste cair uma frigideira numa garrafa*)

<u>"Light" sentences</u>

- You dropped a cotton on a light bulb*
  (*Deixaste cair um algodão numa lâmpada*)
- You dropped a sponge on a tile
  (*Deixaste cair uma esponja num azulejo*)
- You dropped a cork on an iPad
  (*Deixaste cair uma rolha num iPad*)
- You dropped a banknote on a cup
  (*Deixaste cair uma nota numa caneca*)
- You dropped an envelope on a strawberry
  (*Deixaste cair um envelope num morango*)
- You dropped a ribbon on a sushi
  (*Deixaste cair uma fita num sushi*)

**Note*: In original Portuguese language all of the object names consisted of one word.

Appendix B

Results from repeated-measures ANOVA

Repeated-measures ANOVA

Exp.	Omnibus Test	Post-hoc comparisons (segregation of the items by pictures)		
	Sentence by Picture Interaction	Sentence Type	Canonical picture	Non-canonical picture
E.1	$F_1 = 20.91, p < .001$	Heavy	$F_1 = 2.03, p = .16$	$F_1 = 25.26, p < .001$
	$F_2 = 41.48, p < .001$	Light	$F_2 = 1.29, p = .27$	$F_2 = 37.32, p < .001$
E.2	$F_1 = 12.90, p < .001$	Heavy	$F_1 = 1.15, p = .29$	$F_1 = 12.16, p = .001$
	$F_2 = 14.37, p < .001$	Light	$F_2 = 1.08, p = .31$	$F_2 = 9.21, p = .005$
		*Non-action		
E.3	$F_1 = 18.44, p < .001$	Heavy	$F_1 = 3.62, p = .06$	$F_1 = 11.06, p = .001$
	$F_2 = 1.80, p = .19$	Light	$F_2 = 1.18, p = .28$	$F_2 = 0.63, p = .43$
		**Heavy (near)		
		**Light (near)		
E.4	$F_1 = 8.32, p = .005$	Heavy	$F_1 = 0.01, p = .94$	$F_1 = 13.10, p < .001$
	$F_2 = 2.78, p = .10$	Light	$F_2 = 0.08, p = .93$	$F_2 = 4.13, p = .05$
		**Heavy (near)		
		**Light (near)		
E.5	$F_1 = 7.51, p <. 01$	Heavy	$F_1 = 0.11, p = .74$	$F_1 = 8.00, p = .006$
	$F_2 = 0.88, p = .36$	Light	$F_2 = 0.42, p = .84$	$F_2 = 1.25, p = .28$
E.6	$F_1 = 4.18, p = .04$	Heavy	$F_1 = 0.04, p = .85$	$F_1 = 7.62, p < .01$
	$F_2 = 0.71, p = .41$	Light	$F_2 = 0.13, p = .91$	$F_2 = 1.17, p = .29$
E.7	$F_1 = 11.65, p = .001$	Heavy	$F_1 = 0.40, p = .53$	$F_1 = 18.86, p < .001$
	$F_2 = 8.78, p = .007$	Light	$F_2 = 0.26, p = .62$	$F_2 = 10.17, p = .004$

Note. The data were analyzed using the median RTs by condition procedure as in Zwaan et al. (2002). For an accurate comparison of results from ANOVA and linear mixed-effects model (LMEM), it is worth noting that LMEM analysis handles the crossing of F_1 and F_2 simultaneously and takes into account all individual RTs rather than median RTs.

* The reported post-hoc comparisons refer to heavy and light sentence conditions only for the ease of comparison of results across experiments. With regards to the non-action sentence condition, participants verified canonical pictures faster than non-canonical pictures after reading non-action sentences ($F_1 = 11.28$, $p =.001$, $F_2 = 8.92$, $p =.005$).

** The reported post-hoc comparisons refer to heavy and light sentence conditions only for the ease of comparison of results across experiments. Non-canonical pictures were responded to faster after reading the “heavy sentence” with a preposition *on* than the “heavy sentence” with a preposition *near* in the analysis by-participants (Exp.3: $F_1 = 9.18$, $p =.003$, $F_2 = 3.65$, $p =.06$; Exp4: $F_1 = 9.71$, $p =.002$, $F_2 = 2.71$, $p =.109$). No significant difference was observed for canonical pictures involving the “light sentence” with a preposition *on* and the “light sentence” with a preposition *near (*all *Fs < 2).*