COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 44 (2020) e12870

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12870

Is Complex Visual Information Implicated During Language Comprehension? The Case of Cast Shadows

Oleksandr V. Horchak, Margarida Vaz Garrido

Instituto Universitário de Lisboa (ISCTE-IUL), CIS-IUL

Received 28 January 2019; received in revised form 6 May 2020; accepted 25 May 2020

Abstract

Previous research showed that sensorimotor information affects the perception of properties associated with implied perceptual context during language comprehension. Three experiments addressed a novel question of whether perceptual context may contribute to a simulation of information about such out-of-sight objects as cast shadows. In Experiment 1, participants read a sentence that implied a particular shadow cast on a target (blinds vs. an open window) and then verified the picture of the object onto which a shadow was cast. Responses were faster when the shadow of blinds cast on the object matched that implied by the sentence. However, the data did not show the same matching effect for pictures with cast shadows from an open window. In Experiments 2 and 3, we found that verification times for pictures with no cast shadows were faster when preceded by an "open window" sentence, thus suggesting that reading the sentence does not elicit a visual simulation of any specific shadow. Experiment 3 showed that the objects superimposed with a cast shadow of the blinds and blinds themselves were verified faster after reading a "blinds" sentence. However, the results of an order analysis showed the temporal stability of the "blinds shadows" effect, but the disappearance of the "blinds" effect in the second half of the data. We conclude that the results are compatible, to a lesser or greater extent, with multiple accounts, and discuss our findings in the context of a mental imagery view, a mental simulation view, and an amodal representation view.

Keywords: Language comprehension; Cast shadows; Perceptual simulation; Language-induced imagery; Amodal representation

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

1. Introduction

Within the last two decades, a strong body of behavioral (e.g., Horchak, Giger, Cabral, & Pochwatko, 2014; Knoeferle & Guerra, 2016; Lynott & Connell, 2010; Matlock, 2004) and neurophysiological (Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005; Solomon & Barsalou, 2004) evidence emerged supporting the notion that partial re-enactments of sensorimotor states underlie language processing. In this view, understanding a sentence describing someone eating an ice cream on a hot summer day, for example, might require mentally simulating the likely perceptual properties of the object (e.g., a wafer-style ice cream cone), events (e.g., eating), emotions (e.g., pleasure), and a perceptual context (e.g., sunny weather).

Visual simulations constitute one of the strongest types of evidence in favor of the embodied view of language comprehension. In one line of research, Stanfield and Zwaan (2001) instructed participants to read various sentences followed by pictures in either vertical or horizontal orientation and found that responses were faster for picture stimuli that matched the orientation implied by the sentence. A similar study (Zwaan, Stanfield, & Yaxley, 2002) demonstrated that when participants read sentences describing differently shaped items (e.g., a perched eagle vs. a flying eagle), simulation-consistent images were again recognized more quickly than simulation-inconsistent ones. Finally, de Koning, Wassenburg, Bos, and Van der Schoot (2017) and Hoeben Mannaert, Dijkstra, and Zwaan (2017) found that participants also mentally represent object size and object color, respectively. Altogether, these matching effects suggest that comprehenders mentally represent implicit perceptual information on object properties.

Another line of research shows that people may also simulate perceptual contexts while comprehending the language. For instance, Yaxley and Zwaan (2007) asked participants to read sentences like “Through the fogged goggles, the skier could hardly identify the moose” or “Through the clean googles, the skier could easily identify the moose” and later decide whether the pictured object was mentioned in the sentence. Consistent with the simulation account, responses were faster when the resolution of the pictured object and the resolution implied by the sentence matched, thus suggesting that people simulate visibility during language processing. Finally, several reports from eye tracking research revealed that on hearing or reading a description about a specific spatial location, participants tend to direct their eye movements toward those regions of the blank screen that are consistent with the location implied by the linguistic description (Altmann, 2004; Knoeferle & Crocker, 2006; Richardson & Dale, 2005; Spivey & Geng, 2001). Thus, comprehenders, as Barsalou (2003) pointed out, simulate “being there” in the scene and direct their eye movements to where they would look if they were actually observing a real situation.

Bringing the aforementioned literature together, it becomes clear that while reading a sentence people may not only represent the perceptual aspects of a described object but also simulate the implied perceptual context. Such findings lend themselves naturally to the discussion of the constraints on the degree of perceptual simulation underlying sentence processing. On the one hand, previous research indicates that perceptual context may modulate, for instance, the perceptibility of observed referents (Yaxley & Zwaan, 2007). On the other hand, it is unclear whether language comprehension may involve the simulations of objects or aspects of the scene that are "unseen." In this article, we tested this possibility with regard to a visual situation in which the observation of a cast shadow in a scene implies the presence of out-of-sight objects.

Several lines of perception research hint at the possibility that a simulation system could use cast shadows as important cues about the surrounding environment and objects in it. For example, Mamassian, Knill, and Kersten (1998) found that cast shadows not only provide knowledge about the environmental characteristics of the light source but also the shape of casting objects. Braje, Legge, and Kersten (2000) demonstrated that shadows have no effect on naming latency and accuracy during picture recognition. However, the absence or incongruence of a shadow with respect to the casting object may hinder recognition performance (Castiello, 2001). Finally, there is also some evidence suggesting that shadows may even facilitate the visual processing of object shape. As one example, Leek, Davitt, and Cristino (2015) asked participants to first memorize different 3D novel objects with either no shadow, object internal shadow, or both object internal and external (ground) plane shadow and then discriminate previously memorized targets from visually similar distractors. They found that objects learned with their internal cast shadows were recognized faster, as compared to objects learned with no shadow. Thus, it seems reasonable to propose that an interaction between a light source and a caster's intrinsic properties (e.g., shape) might serve as the means to simulate the presence of cast shadows in a described scene.

Combined with previous findings on visual simulation and shadow perception, the experiments presented below were designed to take a first step toward the prediction that language processing may invoke simulation of complex visual scenes involving cast shadows. We operationalized this prediction by asking participants to read sentences like "The sun is shining onto a backpack through the blinds" or "The sun is shining onto a backpack through an open window" and then verify whether the pictured target (i.e., a backpack) appeared in the sentence (see Fig. 1). If participants indeed visually recreate the experience of being in the situation and do not represent a described target in isolation, then their mental representation should contain implicit perceptual information regarding the presence of shadows. Accordingly, participants should be able to anticipate the presence of a cast shadow on a pictured target that corresponds to the one implied by the preceding sentence. This leads to a formulation of a hypothesis that pictorial stimuli should be verified more easily in the "matching shadow" condition and with more difficulty in the "mismatching shadow" condition.

2. Experiment 1

Participants read sentences that implied a particular shadow cast on a target and then verified the picture of the object onto which a shadow was cast. In line with the

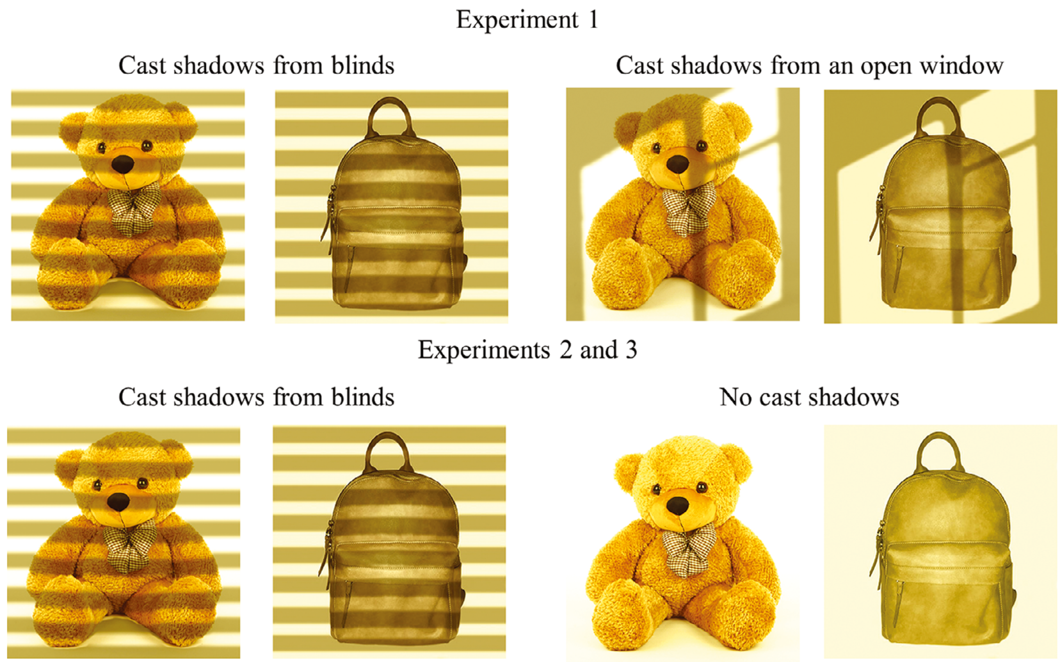

Fig. 1. Samples of pictures depicting objects with cast shadows from blinds (“blinds shadow” condition in Experiments 1–3), cast shadows from an open window (“open window” condition in Experiment 1), and no cast shadows (“open window” condition in Experiments 2 and 3).

hypothesis outlined earlier, responses should be faster when the shadow cast on the pictured target matches the shadow implied by the linguistic description.

2.1. Methods

2.1.1. Participants

Power analysis was conducted in G*Power. By running a power analysis on a repeated measures ANOVA with a medium ($\eta_p^2 = 0.06$) effect size (Faul, Erdfelder, Lang, & Buchner, 2007), we expected to need about 80 participants for each experiment. An estimate of effect size for power analysis is based on the results of previous thematically related research (e.g., Sato, Schafer, & Bergen, 2013; Zwaan et al., 2002). To account for low accuracy scores and compliance with the task requirements, 104 native Portuguese-speaking undergraduate students (M_{age} = 23.23, SD_{age} = 4.50) were recruited as participants in exchange for course credit. Twenty-two participants were male (M_{age} = 24.00, SD_{age} = 3.03) and 82 were female (M_{age} = 23.04, SD_{age} = 4.80). All participants (including those from Experiments 2 and 3) gave informed consent to participate in the study in accordance with the ethical guidelines of the host institution.

2.1.2. Materials

We created 24 experimental sentence pairs and 48 filler sentences (see Appendix A, for samples of all sentences used). The experimental sentence pairs were identical in their form and induced the perspective of a participant (e.g., “You see how the sun is shining onto {an object X} through {the blinds/an open window}”) as we were worried that simulating the presence of a cast shadow in a scene from a different-person perspective could have implications for the mental simulation process (see Dee & Santos, 2011, for a discussion on how the distance or position of a caster impacts shadow perception). Twelve of 48 filler sentences focused on the act of seeing, and the other 36 filler sentences focused on the interaction of participant with an object. Additionally, we created 24 comprehension questions (see Fig. 2) related to filler sentences to keep participants focused on the task as well as enable us to detect the responses of those participants who were not reading for comprehension (i.e., accuracy $< 50\%$).[1] These questions appeared after picture verification and required participants to use the “yes” and “no” keys to respond. Thus, all participants read 24 experimental sentences and 48 filler sentences, as well as responded to 24 comprehension questions. By using twice as many filler sentences (as well as comprehension questions), we hoped to prevent the comprehenders from guessing the purpose of the task.

In addition, 72 images were used to accompany the sentences. Most of the pictures onto which shadow effects were applied came from Google images, but some were created by the first author. The size of each image was set to 385 × 385 pixels. Twenty-four pictures were selected as experimental pairs that depicted the same large-sized referential object onto which a differently shaped shadow was cast. The shape of this shadow either

Filler Sentence	Picture	Comprehension Question
The sun is shining onto a bench through the trees		Was the sun shining onto a bench through the trees? (YES -correct response)
The sun is shining onto a cactus through the wall cracks		Was the sun shining onto a cat through the wall cracks? (NO-correct response)

Fig. 2. Samples of filler sentences, pictures, and comprehension questions.

matched or mismatched the shape of the shadow implied by the experimental sentence. All shadows cast by blinds were horizontal and about the same contrast. We varied, however, the number of stripes shown and their thickness to reduce learning characteristic. All shadows cast by an open window were identical as it was difficult to predict what types of windows (with frames or without frames) participants typically see in their lives. As cast shadows are composed of the light source, the background surface, the viewpoint, and the object casting the shadow, we reasoned that an "open window" sentence may imply the presence of cast shadows in a situation when the casting object is located to the right side of the observer and the light source is coming from above. In this case, one would expect a deformation of the shape of the shadow from an open window similar to that shown in Fig. 1 (see Mamassian et al., 1998, for further information regarding potential sources of a cast shadow displacement). Factors other than shape that contribute to shadow formation and perception, such as light source intensity, and color, were kept approximately constant across all experimental pictures by using a software-based color contrast filter. To create the illusion that the light is cast through blinds or an open window, we used such Adobe Photoshop's functions as displacement map, color balance adjustment layer, levels adjustment layer, and layer mask functions. Finally, 48 pictures were used as filler items. The levels of light source intensity and color applied to filler pictures were approximately the same as those applied to experimental pictures. In addition, 12 of the 48 filler pictures were presented to participants with either a shadow from blinds (six pictures) or a shadow from an open window (six pictures). Thus, participants verified 36 pictures with a shadow cast on a target and 36 pictures with no cast shadows.

2.1.3. Design and procedure

Four lists of stimuli were created and each experimental item appeared in only one condition per list. The conditions in which a sentence–picture pair could occur were as follows: blinds-blinds shadow, open window-open window shadow, blinds-open window shadow, and open window-blinds shadow. Participants were randomly assigned to each list and each participant saw only one list. Therefore, the experiment was a 2 (sentence) × 2 (picture) × 4 (list) design, with sentences and pictures considered as within-participants factors and list considered as a between-participants factor. Note, however, that on reviewer request list was not included as a factor in the statistical analyses because it did not account for a significant portion of the variance in response times (RTs) in all three experiments (Pollatsek & Well, 1995).

Stimulus presentation was delivered through the E-Prime 2.0 software. Participants were instructed to read a sentence, press the Spacebar after understanding it, and finally judge whether the subsequently presented pictured target appeared in the sentence by pressing the "S" button for a "Yes" response and the "N" button for a "No" response. These buttons were chosen to minimize participants' confusion as "S" is the initial for "Sim" (which means "Yes" in Portuguese) response and "N" is the initial for "Não" (which means "No" in Portuguese) response. There were 36 stimuli requiring an "S" response and 36 stimuli requiring an "N" response. Additionally, the instructions warned participants to read sentences attentively to be able to respond to comprehension questions appearing after some pictures. The experiment started with six practice trials to ensure that participants understood the instructions. After each such trial participants saw different feedback screens based on whether the response to the picture was correct or incorrect.

2.1.4. *Data treatment*

Prior to analysis, in all three experiments, filler items, incorrect responses, and trials with absurdly short and long response latencies (shorter than 200 ms and longer than 5,000 ms) were excluded. Second, we examined the distribution of RTs in all of the experimental conditions and found that RTs were positively skewed, thus violating the normality assumption of the general linear model. Consequently, prior to statistical analysis, RTs were log-transformed to reduce positive skew. Third, trials falling outside ±3 *SD*s from the grand condition log-transformed means were removed. Finally, we looked at the distribution of RTs (with Q–Q plots and histograms with normal curve) and found that the data were approximately normally distributed. As a rule of thumb, whenever possible, we aimed to not have more than 5% of the data excluded in each of the experiments (Baayen & Milin, 2010).

2.1.5. *Data analysis*

Accuracy: To test whether accuracy differed across conditions, in all three experiments we first always attempted to use a mixed-effects logistic regression using SPSS (via Generalized Linear Mixed Models Command) in order to test whether two fixed factors, their interaction (sentence type, picture type, sentence by picture interaction), and two random factors (participants and items) predict the probability of a correct response. If the analyses with participants and items included in the model as random intercepts revealed that there was no impact from random effects due to the absence of significant variability in intercepts (i.e., coined as redundant factors in SPSS), the development of a multimodal model was stopped (Heck, Thomas, & Tabata, 2013) and ordinary logistic regression was conducted instead.

Response times: RTs were analyzed using linear mixed-effects models with crossed random effects of participants and items using SPSS. 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) ANOVAS is that this kind of analysis (a) handles the crossing of two random factors simultaneously (Baayen, Davidson, & Bates, 2008) and (b) takes into account individual RTs rather than mean RTs for each participant (Baayen & Milin, 2010). In our mixed model, we used a top-down approach, whereby an attempt was made to (a) fit the full variance–covariance structure of random effects (the so-called maximal model; Barr, Levy, Scheepers, & Tily, 2013) and in case of nonconvergence to (b) use a model selection criterion (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017) based on choosing random effects structure supported by the data.

The model with the most maximal effects structure for the present studies included two main effects and their interaction (sentence type, picture type, sentence by picture interaction) as random slopes and participants and items as random intercepts; maximum likelihood estimation parameter; and an unstructured covariance matrix permitting the covariance between random slopes and random intercepts. Note that in our studies participants gave only one response per individual test item, and hence no slopes by test items are necessary (see Barr et al., 2013, for a discussion). If the model with the most maximal effects structure did not converge, we tested the fit of different covariance and random effects structures with a chi-square likelihood ratio test (−2 log-likelihood in SPSS) and Bayesian information criterion (BIC) based on comparing information criteria of the models. The idea behind these tests is that smaller values mean better fitting models for the data. To allow for check of the transparency of our analyses (e.g., whether the use of a simplified covariance structure or the exclusion of a random effect was justified), we present the results of a number of models containing all fixed factors but different random factors or covariance matrices in Supplementary materials. The best-fitting model selected to report the results is in bold type. Note, however, that in case of convergence of more than one model, we always gave preference to the model that included, at the very least, a random slope for an interaction—the predictor of our major theoretical interest (see Barr et al., 2013, for further discussion).

2.2. *Results and discussion*

2.2.1. *Data trimming*

Sixteen participants were eliminated from the analysis for having accuracy <80% or answering <50% of the comprehension questions correctly. The removal of RTs shorter than 200 ms and longer than 5,000 ms, as well as responses exceeding ±3 *SD*s from the grand condition mean, led to removal of 3.74 % of observations.

2.2.2. *Accuracy*

As can be seen from Table 1, participants' accuracy was high. The analysis of participants' responses using logistic mixed-effects regression with crossed random effects of participants and items (via SPSS Generalized Linear Mixed Models command) showed that the random effects of participants (estimate = 1.332E-7[a]) and items (estimate = 3.634E-8[a]) were redundant parameters for the model. This result suggests that the model was unable to uniquely estimate any variation from participant to participant (or item to item), above and beyond the residual variance from fixed effects to fixed effects. Therefore, logistic regression analysis was employed to predict a probability that a participant would provide a correct response. The predictor variables were sentence type (blinds vs. open window), picture type (blinds shadow vs. open window shadow), and their interaction. Of major interest to our hypothesis, there was a significant interaction between sentence type and picture type.

We segregated participants' responses by pictures to investigate this interaction further. Follow-up analyses revealed that the responses to "blinds shadow" pictures were significantly more accurate when these were preceded by a "blinds" sentence than an "open

Table 1
Accuracy scores (percent correct) and data of logistic regression analysis for Experiment 1

Accuracy Scores		
Picture	Blinds Sentence	Open Window Sentence
Blinds shadow	99%	97%
Window shadow	98%	99%

Logistic Regression Analysis

					95% CI for Odds Ratio	
	B (*SE*)	Wald	Sig.	Odds Ratio	Lower	Upper
Sentence	−3.61 (1.28)	7.98	.005	0.03	0.00	0.33
Picture	−3.20 (1.32)	5.89	.015	0.04	0.00	0.54
Sentence × Picture	2.27 (0.82)	7.65	.006	9.69	1.94	48.44
Constant	9.42 (2.23)	17.86	.000	12292.63		

Note. $R^2 = .01$ (Cox & Snell R Square), .03 (Nagelkerke R Square). Model $\chi^2(3) = 10.36$, $p = .016$.

window" sentence, $\chi^2(1, N = 1{,}056) = 6.49$, $p = .011$. Responses to "open window" pictures, however, were not significantly more accurate when these were preceded by an "open window" sentence than a "blinds" sentence, $\chi^2(1, N = 1{,}056) = 2.61$, $p = .106$. There were also significant effects of sentence type and picture type, but their interpretation is omitted due to the absence of unique slopes for the main effects (i.e., slopes for main effects change based on other variables in the interaction).

2.2.3. *Response times*

The data of major interest are presented in Table 2 and Fig. 3. For the convenience to visualize effects on the original millisecond scale, data are presented using back-transformed mean RTs (Baayen & Milin, 2010). The best converging model included two factors and their interaction as fixed effects (i.e., sentence type and picture type); and a random intercept for participants and items.

The analyses of log-transformed RTs revealed a significant interaction between sentence type and picture type. This interaction was broken down by conducting separate multilevel models on the "blinds shadow" pictures and "open window shadow" pictures. The models specified were the same as the main model but excluded the main effect and interaction term involving the picture type (Field, 2013). The analyses revealed that participants' RTs were positively correlated with corresponding effects in the accuracy rates, thus precluding speed–accuracy tradeoffs. More specifically, the results showed that pictures with cast shadow from blinds were responded to more quickly when preceded by sentences implying a "blinds" shadow than when preceded by sentences implying an "open window" shadow, $b = 0.03$, $t(86.17) = 3.24$, $p = .002$. At the same time, pictures with cast shadow from an open window were responded to equally fast when preceded by both "blinds" and "open window" sentences, $b = 0.00$, $t(89.43) = 0.34$, $p = .737$.

Table 2
Mean back-transformed latencies of correct responses (in ms) with 95% confidence intervals and parameter estimates of log-transformed RTs for Experiment 1

Response Times

Picture	Blinds Sentence, *M* [95% CI]	Open Window Sentence, *M* [95% CI]
Blinds shadow	733 [707, 760]	790 [759, 823]
Window shadow	749 [721,778]	756 [728, 785]

Parameter Estimates

				95% CI	
	B (*SE*)	*t*	Sig.	Lower	Upper
Intercept	2.79 (0.04)	79.66	.000	2.72	2.86
Sentence	0.07 (0.02)	3.23	.001	0.03	0.11
Picture	0.04 (0.02)	2.12	.034	0.00	0.08
Sentence × Picture	−0.03 (0.01)	−2.43	.015	−0.06	−0.01

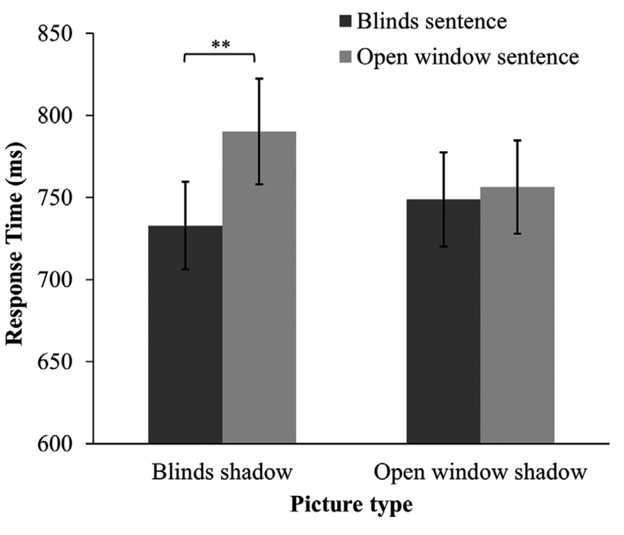

Fig. 3. Mean back-transformed response times (in milliseconds) and confidence intervals for pictures with cast shadows from blinds and cast shadows from an open window after these were preceded by a "blinds" sentence or an "open window" sentence in Experiment 1. $**p < .01$.

Given the significant interaction term, the effects of sentence type and picture type from Table 3 are not interpreted due to the absence of unique slopes for the main effects (i.e., slopes change based on other variables in the interaction).

These results are only partially consistent with our prediction and suggest that participants represented the implied shadow from blinds but not the shadow from an open

Table 3
Accuracy scores (percent correct) and data of logistic regression analysis for Experiment 2

Accuracy Scores

Picture	Blinds Sentence	Open Window Sentence
Blinds shadow	98%	98%
No shadow	98%	98%

Logistic Regression Analysis

	B (*SE*)	Wald	Sig.	Odds Ratio	95% CI for Odds Ratio Lower	95% CI for Odds Ratio Upper
Sentence	−0.09 (1.02)	0.01	.934	0.92	0.12	6.81
Picture	0.24 (1.07)	0.05	.823	1.27	0.16	10.36
Sentence × Picture	−0.12 (0.65)	0.03	.854	0.89	0.25	3.19
Constant	3.97 (1.67)	5.67	.017	53.12		

Note. R^2 = .00 (Cox & Snell R Square), .00 (Nagelkerke R Square). Model $\chi^2(3) = 0.72$, $p = .870$.

window. However, such an interpretation is only valid if one assumes the following two scenarios. The first is that our picture of a cast shadow from an open window may lack shared content with a simulation of a sentence about cast light from an open window, precisely because participants might have accumulated different situated conceptualizations associated with using an open window. If, for example, one participant has consistently observed a situation in which the light is cast onto an object through a double hung window, whereas another participant experienced how the light is cast through a casement window, then their anticipatory inferences about the caster's shadow would be different (see Barsalou, 2015, for further discussion). The second possibility is that perceptual simulation of cast light from an open window simply does not presuppose the existence of any shadows. This assumption is not counterintuitive because if one simulates a described object as being placed close to a window opening, then the occlusion of light sources may indeed seem unlikely. Teasing apart these two possibilities strikes us as impossible as it would require collecting data from only those participants who always see the same types of windows (e.g., with frame or without frame) in their lives. However, if comprehending how the sun is shining onto an object through an open window does not involve simulating a shadow of specific window shape, then we should see a statistical difference between matching and mismatching conditions when participants verify the pictures of objects onto which no shadows are cast, as compared to when they verify the pictures of objects with cast shadows from blinds. This difference would be compatible with either facilitation in the matching sentence condition (i.e., pictures with no shadows are responded to more quickly because they are preceded by "open window" sentences) or interference in the mismatching sentence condition (i.e., pictures with no shadows are responded to more slowly because they are preceded by "blinds" sentences). Experiment 2 was designed to provide such evidence.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Ninety-three native Portuguese-speaking undergraduate students (M_{age} = 21.53, SD_{age} = 4.61) participated in the experiment in exchange for course credit. Twenty-two participants were male (M_{age} = 22.60, SD_{age} = 4.63) and 71 were female (M_{age} = 21.23, SD_{age} = 4.59).

3.1.2. Materials

All the sentences were the same as in Experiment 1. The experimental pictures were the same, except that "open window" shadows were removed (Fig. 1). That is, participants now verified the pictures depicting cast light in the following two conditions: with a shadow of blinds cast on a pictured target ("blinds" sentence condition) and without any shadow cast on a pictured target ("open window" sentence condition). The filler pictures were also the same, except that 24 of the 48 pictures were presented with a cast shadow from blinds. Thus, participants verified 36 images with a "blinds" shadow and 36 images without any shadow.

3.1.3. Design and procedure

The design and procedure were the same as for Experiment 1.

3.2. Results and discussion

3.2.1. Data trimming

The screening procedures from Experiment 1 were repeated: Nine participants were removed for having accuracy less than 80% or answering less than 50% of comprehension questions correctly. The removal of responses shorter than 200 ms and longer than 5,000 ms, as well as responses exceeding ±3 *SD*s from the grand condition mean led to removal of 6.85% of the observations. A removal of more than 5% of the observations (our desired threshold) was mainly due to the presence of five participants with unusually high RTs. Note, however, that although these participants had about 50%–75% of valid observations per condition, their data had no effect on the performed statistical analyses (i.e., a model with the most maximal effects structure that converged was the same with or without the data from these participants).

3.2.2. Accuracy

As it can be seen from Table 3, participants' accuracy was high. The analysis of participants' responses using logistic mixed-effects regression with crossed random effects of participants and items (via SPSS Generalized Linear Mixed Models command) showed that the random effects of participants (estimate = 9.169E-8[a]) and items (estimate =

2.040E-7[a]) were redundant parameters for the model. This result suggests that the model was unable to uniquely estimate any variation from participant to participant (or item to item), above and beyond the residual variance from fixed effects to fixed effects. Therefore, logistic regression analysis was employed to predict a probability that a participant would provide a correct response. This analysis revealed that none of the main predictors (i.e., sentence, picture, or their interaction) was significant.

3.2.3. Response times

The data of major interest are presented in Table 4 and Fig. 4. The model with the most maximal effects structure that converged included all fixed effects as before, participants and items as random intercepts, and a sentence by picture interaction as a random slope.

The analyses of log-transformed RTs revealed a significant crossover interaction between sentences and pictures. This interaction was broken down by conducting separate multilevel models on the "blinds shadow" pictures and "no shadow" pictures. The models specified were the same as the main model but excluded the main effect and interaction term involving the picture type. The data showed that pictures with superimposed cast shadows of the blinds were responded to more quickly when preceded by a "blinds" sentence than when preceded by an "open window" sentence, $b = 0.04$, $t(83.37) = 3.18$, $p = .002$. Similarly, pictures with no superimposition were responded to more quickly when preceded by an "open window" sentence than when preceded by a "blinds" sentence, $b = -0.03$, $t(84.63) = -2.96$, $p = .004$. These results, aside from replicating those from Experiment 1 for the "blinds shadow" stimuli, also support our hypothesis that reading how the sun is shining onto an object through an open window does not lead to a simulation of a shadow of specific window shape.

Table 4
Mean back-transformed latencies of correct responses (in ms) with 95% confidence intervals and parameter estimates of log-transformed RTs for Experiment 2

Response Times

Picture	Blinds Sentence, *M* [95% CI]	Open Window Sentence, *M* [95% CI]
Blinds shadow	716 [692, 741]	778 [749, 808]
No shadow	804 [774, 834]	751 [725, 778]

Parameter Estimates

				95% CI	
	B (*SE*)	*t*	Sig.	Lower	Upper
Intercept	2.71 (0.03)	79.81	.000	2.64	2.78
Sentence	0.10 (0.02)	4.88	.000	0.06	0.14
Picture	0.11 (0.02)	5.61	.000	0.07	0.15
Sentence × Picture	−0.06 (0.01)	−5.02	.000	−0.09	−0.04

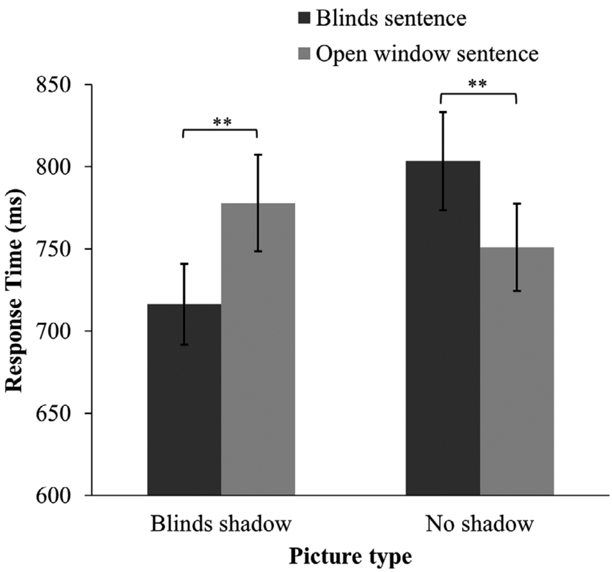

Fig. 4. Mean back-transformed response times (in milliseconds) and confidence intervals for pictures with cast shadows from blinds and no cast shadows after these were preceded by a “blinds” sentence or an “open window” sentence in Experiment 2. $^{**}p < .01$.

3.3. Comparison of the RT effects between Experiments 1 and 2 for “open window” stimuli

Upon observing the difference in results between Experiment 1 and Experiment 2, we decided to run linear mixed-effects analyses on the combined data from Experiments 1 and 2 as before, but now with experiment as a between-participants factor. Given that Sentence × Picture interaction regarding open window stimuli was significant in Experiment 2 and not in Experiment 1, we expected to find some statistical difference between the results of Experiments 1 and 2. Specifically, we wanted to test whether the effect for “open window” stimuli in Experiment 2 was significantly greater than the lack of effect for “open window” stimuli in Experiment 1. To this end, we split the data file by pictures and ran a linear mixed-effects model whose best effects structure that converged included fixed effects of sentence, experiment, and their interaction; random slope for an interaction between experiment and sentence; and a random intercept for participants and items. The results of a three-way interaction between experiment, picture type, and sentence type showed that the difference in results between fully open window and open window with frames was significant, $b = -0.03$, $t(218.77) = -2.45$, $p = .015$.

4. Experiment 3

A crossover interaction between pictures and sentences in Experiment 2 is consistent with our hypothesis that language comprehension involves simulating cast shadows, but does not compel it. First, it is unclear whether the reported data reflect language-induced imagery rather than mental simulation, given that each experimental sentence in Experiments 1 and 2 began with "You see how⋯." Relatedly, it is possible that while reading the sentences participants were not simulating a complex scene of light passing through blinds and casting shadows on the specified object, but rather that they were simply simulating the blinds and the object. In other words, the simulation of the blinds themselves could be enough to facilitate the processing of visually similar objects with horizontal stripes.

To address the first concern, we removed the "You see how" part from the target sentences and asked participants to read sentences like "The sun is shining onto a backpack through {the blinds/an open window}." Our prediction remained unchanged: Participants should be faster to indicate whether the object was mentioned in the previously read sentence when the shadow cast on the perceived target matches the one implied in the sentence. To address the second concern, we added a picture condition where blinds themselves (not their cast shadows) were superimposed on the pictures of the objects. Note that these new pictures were identical to those with a cast shadow from blinds in terms of the number (and position) of stripes shown (see Fig. 5). If reading the sentence elicits a visual simulation of the cast shadow itself, then participants should not be faster at identifying objects with superimposed blinds, as compared to objects shown with a cast shadow of the blinds. Thus, relative to Experiment 2, two major things changed for Experiment 3: the sentence frame and addition of blinds pictures.

4.1. Methods

4.1.1. Participants

One hundred and twenty-two native Portuguese-speaking undergraduate students ($M_{age} = 19.87$, $SD_{age} = 2.53$) participated in the experiment in exchange for course credit. Twenty-two participants were male ($M_{age} = 20.23$, $SD_{age} = 1.97$) and 100 were female ($M_{age} = 19.79$, $SD_{age} = 2.64$).

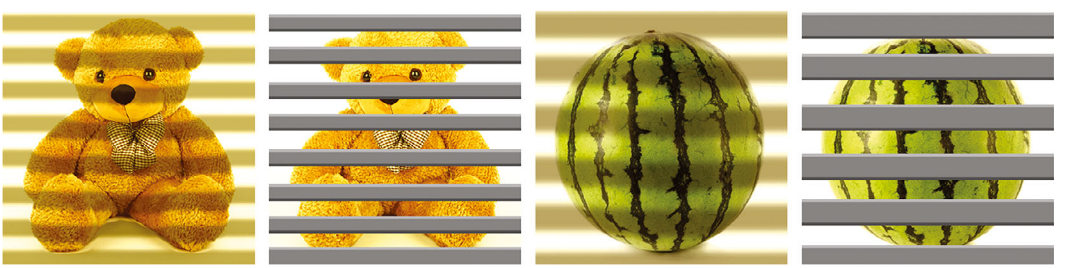

Fig. 5. Samples of pictures when cast shadows from blinds (left) and blinds themselves (right) are superimposed on the pictures of the objects.

4.1.2. Materials

The experimental stimuli from Experiment 2 were used except that the part "You see how" was removed from the sentences. Other important differences were as follows. First, 12 new experimental sentence pairs were constructed that matched the form of other experimental sentences (e.g., "The sun is shining onto {an object X} through {the blinds/an open window}"). Second, 12 new experimental pictures with a shadow of blinds cast on a pictured target and without any shadow cast on a pictured target were created. Third, 36 new pictures were designed in which blinds themselves were superimposed on the objects (see Fig. 2). Fourth, to further disguise experimental items, 12 new sentence–-picture pairs depicting an object (onto which a blinds shadow was cast) that mismatched the one mentioned in the sentence were used. All other filler items were the same as those from Experiment 2 except for one difference: Only 12 of 48 pictures were presented with a blinds shadow. Thus, participants verified 48 images in which blinds or their cast shadows were superimposed on the pictures of the objects and 48 images with no superimposition.

4.1.3. Design and procedure

To counterbalance items and conditions, six lists of stimuli were created. The conditions in which a sentence–picture pair could occur were as follows: blinds-blinds shadow, blinds-blinds, blinds-no superimposition, open window-blinds shadow, open window-blinds, open window-no superimposition. Thus, the experiment was a 2 (sentence: blinds/open window) × 3 (picture: blinds shadow/blinds/no superimposition) × 6 (list) design, with sentences and pictures considered as within-participants factors and list considered as a between-participants variable. The procedure was identical to that in Experiment 2.

4.2. Results and discussion

4.2.1. Data trimming

The screening procedures from Experiment 2 were repeated: Data from 16 participants who responded correctly to <80% of experimental items or <50% of comprehension questions were eliminated. Data from one additional participant had to be removed because of having only one valid response in two experimental conditions. The removal of responses shorter than 200 ms and longer than 5,000 ms, as well as responses exceeding ±3 *SD* from the grand condition mean, led to removal of 3.97 % of observations.

4.2.2. Accuracy

As it can be seen from Table 5, participants' accuracy was high. The analysis of participants' responses using logistic mixed-effects regression with crossed random effects of participants and items (via SPSS Generalized Linear Mixed Models command) showed that the random effects of participants (estimate = 8.091E-7[a]) and items (estimate = 1.896E-7[a]) were redundant parameters for the model. This result suggests that the model was unable to uniquely estimate any variation from participant to participant (or item to

Table 5
Accuracy scores (percent correct) and data of logistic regression analysis for Experiment 3

Accuracy Scores

Picture	Blinds Sentence	Open Window Sentence
Blinds shadow	99%	96%
No shadow	97%	96%
Blinds	97%	94%

Logistic Regression Analysis

					95% CI for Odds Ratio	
	B (*SE*)	Wald	Sig.	Odds Ratio	Lower	Upper
Sentence	−1.21 (0.49)	6.23	.013	0.30	0.12	0.77
Picture	−0.37 (0.39)	0.93	.335	0.69	0.32	1.47
Sentence × Picture	0.31 (0.23)	1.83	.176	1.36	0.87	2.13
Constant	5.01 (0.84)	35.33	.000	150.51		

Note. R^2 = .00 (Cox & Snell R Square), .02 (Nagelkerke R Square). Model $\chi^2(3) = 14.76$, $p = .002$.

item), above and beyond the residual variance from fixed effects to fixed effects. Therefore, logistic regression analysis was employed to predict a probability that a participant would provide a correct response. This analysis revealed that sentence type was a significant predictor, reflecting the fact that the odds of producing a correct response were lower when participants read "open window" sentences than when they read "blinds" sentences.

4.2.3. Response times

The data of major interest are presented in Table 6 and Fig. 6. There was a successful convergence of the model with two fixed factors and their interaction as fixed effects; participants and items as random intercepts; and picture, and sentence by picture interaction as random slopes. The interaction between sentence type and picture type was significant. This interaction was broken down by conducting separate multilevel models on the "blinds shadow" pictures, "blinds" pictures, and "no shadow" pictures. The models specified were the same as the main model but excluded the main effect and interaction term involving the picture type. Note, however, that due to convergence problems the models for "blinds shadow" and "blinds" stimuli had to be simplified. More specifically, for the "blinds shadow" and "blinds" stimuli we removed a random slope for a sentence (see Supplementary materials, for more details).

The results showed that pictures with superimposed cast shadows of the blinds were responded to more quickly when preceded by a "blinds" sentence than when preceded by an "open window" sentence, $b = 0.04$, $t(1045.82) = 4.93$, $p < .001$. Similarly, pictures with no superimposition were responded to more quickly when preceded by an "open window" sentence than when preceded by a "blinds" sentence, $b = -0.03$, $t(97.94) = -3.19$, $p = .002$. Finally, participants were also significantly faster to verify

Table 6
Mean back-transformed latencies of correct responses (in ms) with 95% confidence intervals and parameter estimates of log-transformed RTs for Experiment 3

Response Times

Picture	Blinds Sentence, *M* [95% CI]	Open Window Sentence, *M* [95% CI]
Blinds shadow	649 [627, 671]	707 [683, 732]
No shadow	696 [672, 722]	653 [630, 678]
Blinds	738 [711, 767]	765 [737, 793]

Parameter Estimates

				95% CI	
	B (*SE*)	*t*	Sig.	Lower	Upper
Intercept	2.81 (0.02)	116.64	.000	2.77	2.86
Sentence	0.06 (0.01)	4.48	.000	0.03	0.09
Picture	0.01 (0.01)	1.23	.219	−0.01	0.03
Sentence × Picture	−0.02 (0.01)	−3.99	.000	−0.04	−0.01

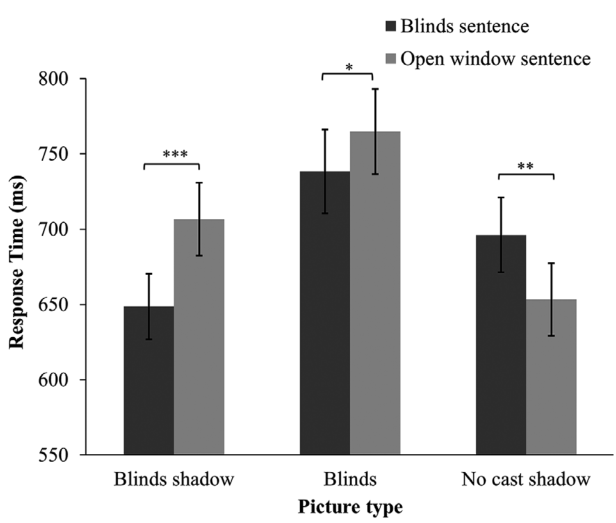

Fig. 6. Mean back-transformed response times (in milliseconds) and confidence intervals for pictures with cast shadows from blinds, blinds themselves, and no cast shadows after these were preceded by a "blinds" sentence or an "open window" sentence in Experiment 3. *$p < .05$, ***$p < .001$, **$p < .01$, *$p < .05$.

"blinds" pictures when these were preceded by a "blinds" sentence than by an "open window" sentence, $b = 0.02$, $t(1012.34) = 2.04$, $p = .042$. Finally, we segregated the items by sentences to test whether there was a statistical difference in participants' responses for "blinds shadow" and "blinds" pictures after reading a "blinds" sentence. The model with the most maximal effects structure that converged (fixed factor for picture, an intercept for participants and items) showed that "blinds shadow" pictures were verified significantly faster than "blinds" pictures after reading a "blinds" sentence, $b = -0.06$, $t(1051.53) = -7.16$, $p < .001$. Presumably this last finding suggests that blinds themselves were not at the focus of participants' attention (but rather an object behind them) while reading the target sentences—a scenario that is more compatible with a simulation of a cast shadow rather than a caster itself.

As the above pattern of results did not allow us to clearly distinguish between (a) participants simulating blinds or (b) participants simulating blinds' shadows, we ran an order analysis for the effect of sentence type on picture verification by looking at first-half versus second-half data. Specifically, we were interested to know if the observed compatibility effects for "blinds", "blinds shadow," and "no shadow" pictures were different or about constant throughout the study.

The results of major interested are presented in Table 7. With regard to the first half of the experiment, logistic regression analysis showed that sentence type ($b = -0.25$, $SE = 0.65$, $p = .700$), picture type ($b = 0.62$, $SE = 0.56$, $p = .274$), and their interaction ($b = -0.27$, $SE = 0.32$, $p = .394$) were not significant predictors of a correct response.

The results of the best converging linear mixed-effects model (included random intercepts for participants and items and a random slope for a sentence by picture interaction) revealed that the interaction between sentence type and picture type was significant in the first half of the experiment, $b = -0.03$, $t(1579.43) = -3.15$, $p = .002$. The segregation of the items by pictures showed that pictures with superimposed cast shadows of the blinds were responded to more quickly when preceded by a "blinds" sentence than when preceded by an "open window" sentence, $b = 0.04$, $t(460.76) = 2.98$, $p = .003$. Similarly, pictures with no superimposition were responded to more quickly when preceded by an "open window" sentence than when preceded by a "blinds" sentence, $b = -0.03$, $t(93.73) = -2.46$, $p = .016$. Finally, participants also tended to verify "blinds" pictures

Table 7
Accuracy scores (percent correct) and mean back-transformed response times (in milliseconds) for the first half and the second half of Experiment 3

	Sentence Condition			
	Blinds		Open Window	
Picture Condition	ACC	RT	ACC	RT
First half of the experiment				
Blinds shadow	99%	689	96%	743
No shadow	98%	739	95%	695
Blinds	96%	796	94%	845
Second half of the experiment				
Blinds shadow	99%	612	96%	675
No shadow	96%	656	98%	616
Blinds	98%	689	94%	698

faster when these were preceded by a “blinds” sentence than by an “open window” sentence, $b = 0.03$, $t(95.21) = 1.92$, $p = .058$.

With regard to the second half of the experiment, logistic regression analysis showed that sentence type ($b = -2.33$, $SE = 0.76$, $p = .002$), picture type ($b = -1.41$, $SE = 0.58$, $p = .014$), and their interaction ($b = 0.95$, $SE = 0.34$, $p = .006$) were significant predictors of a correct response. The segregation of participants’ responses by pictures revealed that the above significance of the interaction is attributable to the fact that “blinds shadow” pictures ($\chi^2(1, N = 630) = 5.53$, $p = .019$) and “blinds” pictures ($\chi^2(1, N = 630) = 5.04$, $p = .025$) were responded to significantly more accurately when preceded by a “blinds” sentence than an “open window” sentence. The results of the maximal linear mixed-effects model (included random intercepts for participants and items and a random slope for sentence, picture, and a sentence by picture interaction) revealed that the interaction between sentence type and picture type was significant in the second half of the experiment, $b = -0.02$, $t(97.88) = -2.40$, $p = .018$. The segregation of the items by pictures showed that pictures with superimposed cast shadows of the blinds were responded to more quickly when preceded by a “blinds” sentence than when preceded by an “open window” sentence, $b = 0.04$, $t(484.59) = 4.11$, $p < .001$. Similarly, pictures with no superimposition were responded to more quickly when preceded by an “open window” sentence than when preceded by a “blinds” sentence, $b = -0.03$, $t(97.86) = -2.58$, $p = .011$. However, in contrast to first-half data, participants verified “blinds” pictures almost equally fast when these were preceded by a “blinds” sentence and an “open window” sentence, $b = 0.01$, $t(97.36) = 0.64$, $p = .521$. Thus, these results suggest that the simulation effect stayed about constant throughout Experiment 3 for the “blinds shadow” pictures and “no shadow” pictures, but not for the “blinds” pictures.

4.3. Comparison of results between “blinds shadow” and “blinds” stimuli

To further investigate whether participants were simulating blinds or cast shadow of blinds, we conducted logistic regression and linear mixed-effects analyses as before, but this time excluded the pictures pertaining to an “open window” condition from the analysis in order to test directly whether there were any main effects or interaction between two critical pairs: picture type (blinds shadow vs. blinds) and sentence type (blinds vs. open window).

Logistic regression analyses showed that sentence type ($b = 0.1$, $SE = 0.72$, $p = .991$), picture type ($b = 1.69$, $SE = 0.93$, $p = .068$), and their interaction ($b = -0.68$, $SE = 0.52$, $p = .192$) were not significant predictors of a correct response. The data from RT analysis, however, are less conclusive. The model with the most maximal effects structure that converged (fixed effects and their interaction; random intercept for participants and items; and random slopes for sentence, sentence by picture interaction) showed that the interaction between sentence type and picture type did not reach statistical significance, $b = 0.02$, $t(1927.42) = 1.95$, $p = .051$. There was, however, a statistically strong main effect of picture type ($b = -0.08$, $t(2141.33) = -4.49$, $p < .001$), reflecting the fact that “blinds” pictures were responded to more slowly than “blinds shadow” pictures, regardless of sentence condition. Thus, when put alongside evidence from the first half versus second half of the experiment, it appears that there are two possible interpretations (and not necessarily mutually exclusive) of the results: (a) Participants were simulating cast shadows; and (b) participants were simulating the objects and the blinds when they read experimental sentences.

5. General discussion

The current research considered the question of whether a simulation system invokes information from cast shadows during language comprehension. Experiment 1 provided only partial support for this prediction as we found evidence for visual simulation on sentences that implied cast shadows from blinds and not an open window. Experiment 2 examined the hypothesis that comprehending how the sun is shining onto an object through an open window does not involve simulating a shadow of specific window shape. In accordance with this hypothesis, we found that verification times for pictures with no superimposition of cast shadow from an open window were faster when they were preceded by an "open window" sentence than when preceded by a "blinds" sentence. Furthermore, we replicated (both in Experiments 2 and 3) the results from Experiment 1 in that participants were again faster in their responses when the shadow of the blinds cast on the pictured target matched that implied by the sentence. Experiment 3 investigated whether the results from Experiments 1 and 2 are better accounted for by a mechanism of perceptual simulation or a mechanism of mental imagery. The results showed that in a complex scene of light passing through blinds, participants were mentally representing both the (a) blinds and (b) blinds' shadows, thus suggesting that the results are consistent with both of the aforementioned mechanisms.

As far as our materials are concerned, one could argue that our "blinds" stimuli may be consistent with "open window" stimuli in situations when the light is shining onto a target object through the blinds that are lowered over an open window. In other words, the meaning of light coming through an open window is not necessarily incompatible with the meaning of light coming through the blinds. However, if it were really the case that two different experimental sentences (blinds vs. open window) led to an almost identical perceptual simulation of the situation, then there should have been no differences in participants' verification of pictures. Recall that just as seeing a shadow from blinds speeded up participants' recognition of the target object after reading a "blinds" sentence (as compared to an "open window" sentence), so too not seeing a shadow from blinds speeded up participants' recognition of the target object after reading an "open window" sentence (as compared to a "blinds" sentence).

As far as our interpretation of results is concerned, we conclude that multiple accounts, to a lesser or greater extent, are in line with the observed pattern of results. The first interpretation is more in line with a language-induced imagery account (e.g., Bergen, Lindsay, Matlock, & Narayanan, 2007), which suggests that participants were simulating the objects and the blinds (not the blinds' shadows) while reading the target sentences.[2]

On this account, when participants read the word “blinds,” they visually simulated blinds themselves that just happen to be visually similar to cast shadows from blinds. In other words, the simulation of actual blinds was enough to facilitate processing of objects with horizontal stripes. Indeed, the results from Experiment 3 provide support for this conclusion as they demonstrated a significant compatibility effect in the “blinds pictures” condition. Furthermore, the critical interaction between blinds and blindsshadows in producing a compatibility effect (RT difference for compatible vs. incompatible sentences) was marginal but still not significant ($p = .051$).

The second interpretation is more in line with an account of complex mental simulation, suggesting that participants simulated cast shadows when reading the sentences. Two aspects of our results are worth noting in this regard. First, it is important to remember that participants’ task was not to decide whether a picture showing blinds in front of objects is more consistent with the sentence than the one with shadows from blinds. Participants’ task was to merely decide whether any of the pictured objects was mentioned in the sentence. Hence, if reading the “blinds” sentence led to the perceptual simulation of actual blinds, then verification times for “blinds” pictures should have been, at the very least, just as fast as verification times for “blinds shadow” pictures, precisely because the word “blinds” was mentioned in the sentence. The fact that the visual occlusion caused by the blinds led to an overall increased reaction times in the “blinds picture” condition may be interpreted as an indication of the fact that the focus of participants’ attention during sentence processing were target objects (e.g., a teddy bear) and not the “secondary” objects (i.e., blinds) occluding the light source. Second, the simulation effect for “blinds shadows” pictures (as well as “open window” pictures) stayed about constant throughout the experiment, but the marginally significant effect for “blinds” pictures appeared ($p = .057$) in the first half of the experiment and vanished in the second half of the experiment ($p = .521$).

Although the results from first-half vs. second-half data showed that the matching effect in the “blinds shadows” condition does not require time to build up, at this point we cannot determine why the “blinds” effect decreased over time. Indeed, if the process of simulating actual blinds is not something that happens in a real-life language comprehension scenario, then it is more reasonable to expect that the effect should increase over the course of the experiment, rather than decrease. There are two possibilities for why this happened. One possibility suggests that on reading the sentence “The sun is shining onto a teddy bear through the blinds” comprehenders initially simulated blinds themselves (or some mixture of both blinds and blinds shadows), but on fully reenacting neural states that represent how blinds look like in a situation that contains a light source, a target, and a direction of light motion, comprehenders primarily simulated cast shadows from blinds (as evidenced by second-half data). However, such a scenario seems more plausible at the level of a single trial (i.e., in the course of a few seconds), but not, as we demonstrated, at the level of the whole experiment (i.e., in the course of many minutes). Another possibility suggests the presence of a contrast effect (see Williams, 1983, for more information), suggesting that verification times of “blinds” pictures were affected due to prior exposure to “blinds shadow” pictures that are related on the “blinds” dimension. Specifically, within the time course of an experiment, participants could have found objects superimposed with blinds themselves less relevant or appealing, precisely because the response was made in the presence (when taken experiment as a whole) of objects superimposed with cast shadows from blinds rather than in isolation. Such a possibility is supported, in part, by previous research on contrast effects in judgments of object size. For example, there is evidence that when participants are asked to judge the size of two central circles of the same size, the judgment of a central circle's size is biased in the presence of other surrounding circles (Schiffman, 1976). Thus, a definitive response to this question must await further empirical research.

Much attention in the present paper was devoted to the accounts of automated perceptual simulation and language-induced mental imagery. However, it is worthwhile to note that the results from a sentence–picture verification paradigm alone are not sufficient to determine the extent to which the visual system is recruited during language comprehension (see Ostarek & Huettig, 2019, for a detailed discussion). Indeed, there appears to be no direct experimental support for simulation-based accounts in the sentence–picture task at this point. Furthermore, recent evidence contests the generalizability of the findings obtained with this paradigm. For example, a study of Ostarek, Joosen, Ishag, de Nijs, and Huettig (2019) tested with visual noise whether mental representation of object shape relies on perceptual simulation in the sentence–picture verification task (by using the materials from the study of Zwaan et al., 2002) and found no evidence that visual processes were recruited during the task. Therefore, on the basis of collected data we cannot rule out the possibility that our findings may also be compatible, at least in part, with amodal-based accounts of cognition, which suggest that sensorimotor information, though it contributes to cognitive performance, is generally not central for language comprehension (see, e.g., the grounding-by-interaction model by Mahon & Caramazza, 2008). However, it would be fair to mention in this context that relegating sensorimotor information an "ornamental" role is contested by evidence from other methods, such as electroencephalography (Coppens, Gootjes, & Zwaan, 2012), neurophysiology (Glenberg et al., 2008), judgment tasks (Horchak, Giger, & Garrido, 2016), continuous flash suppression (Edmiston & Lupyan, 2017; Ostarek & Huettig, 2017), and eye tracking (Speed & Vigliocco, 2014).

Thus, an important qualification of the present research is that our results do not constitute direct evidence for the claim that language comprehension relies on perceptual simulation and/or mental imagery, though our findings are compatible with such a claim. Still, we believe that the use of a sentence–picture verification paradigm is well-justified for the first step of any new research program like ours as it is capable of demonstrating whether there is, indeed, a finding worthy of further scientific inquiry. Note that relative to the strong case for the activation of information regarding different object properties (e.g., orientation, shape, etc.), the case for the integration of object representations with environmental context is growing. Admittedly, our results do not provide a strong support for the hypothesis that complex visual information is simulated during language comprehension. Nonetheless, our data do shed light on the conditions that may define the integration of such complex visual information into a mental or situation model (e.g., Johnson-Laird, 1983; Kintsch & van Dijk, 1978; Zwaan, Magliano, & Graesser, 1995), thus suggesting that there are still many aspects of the language-perception interplay that future research needs to address. As one example, congruency effects in our study were only limited to the shadows cast by blinds (and not an open window). Thus, it is interesting to speculate on the conditions that might require the simulation of described perceptual information. Our prediction is that perceptual simulation may play a role in the mental representation of cast shadows only when these sufficiently resemble the caster, or the prototype (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Thus, if a caster (e.g., blinds) is easily integratable with its shadow, then initially activated linguistic forms could give rise to a simulation of a more complex scene involving the light source. If, however, a caster (e.g., a window) is not as easily integratable with its shadow, then the simulation of perceptual context will be considered of negligible benefit and symbolic processing will reign supreme. At the theoretical level, this means that comprehenders may engage either linguistic or embodied processing (or mixtures of the two systems) depending on the extent to which the shadow and its caster distinguish from one another. At a more empirical level, this suggests that without considering each shadow type (e.g., prototypical vs. non-prototypical) as unique one could arrive at the woefully erroneous conclusion about the specific mental mechanisms underlying the integration of implied "cast shadow" information in sentence processing.

As another example, one may ask if the observed pattern of results might change if another critical aspect of shadow perception is varied, such as a light source. Recent perception research hints at such a possibility as there is, for example, empirical evidence that the 7- to 8-month-old infants perceive an object's lightness in shadows and are able to detect the differences between natural and unnatural lightness changes (Sato, Kanazawa, & Yamaguchi, 2017). Furthermore, there are even findings suggesting that eye pupils adjust to imaginary light (Laeng & Sulutvedt, 2014). Thus, the challenge for future research is to examine in more detail the theoretical ramifications regarding the complexity involved in mental representation of the visual world. In our opinion, future studies in which both caster information (e.g., shape, size, or position) and light information (e.g., color, intensity) are varied across sentence and picture stimuli would be of great theoretical interest to provide some clues as to exactly how and when the light source contributes to a representation of cast shadows of specific shape, size, location, or intensity. This would, undoubtedly, shed some further light on the specific parameters and boundary conditions of the effect and, more generally, on the importance of perceptual simulation for successful language comprehension. Finally, future research should consider using more sophisticated paradigms that have been developed to directly test the involvement of visual processes in language comprehension and conceptual processing (e.g., see Montero-Melis, Isaksson, van Paridon, & Ostarek, 2019; Tsuchiya & Koch, 2005, for a discussion of a continuous flash suppression paradigm). In addition, modified versions of a sentence–picture verification task can be used,[3] where an assumed simulation is interfered by a concurrent perceptual task and its effect on the picture verification task is assessed (e.g., see Ostarek et al., 2019, for an example of such a paradigm where visual noise is used).

In conclusion, our data join with Yaxley and Zwaan (2007) results to demonstrate that perceptual context is not represented as separate from a referential object in perception. Furthermore, our data are consistent with other findings showing the role of sensorimotor simulations in language comprehension (e.g., Borreggine & Kaschak, 2006; Horton & Rapp, 2003; Richardson & Matlock, 2007; Wilson & Gibbs, 2007). Yet our results go beyond these previous findings by providing some of the first evidence that as a person comprehends a sentence, he or she might not only represent objects within a visual field, but also represent out-of-sight objects by means of the interaction between the light environment and the caster that occludes the primary light source. The challenge for future research is to establish which theoretical framework provides a better account of such findings.

Acknowledgments

This research was supported by the Foundation of Science and Technology of Portugal by grants awarded to the first (SFRH/BPD/115533/2016) and second (PTDC/MHC-PCN/5217/2014) authors. We thank anonymous reviewers for helpful comments on previous versions of this article.

Conflict of Interest

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.

Author Contributions

OH idealized the study, designed the picture stimuli, and drafted the manuscript. OH and MG designed the sentence stimuli, involved in acquisition of data, analysis and interpretation of data, critical revision of the manuscript, and approval of the submitted version for publication.

Open Research badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://osf.io/67gyx/?view_only=c12ef7a2bcc6472d9577b88da2f967c1.

Notes

1. Participants' accuracy when responding to the comprehension questions was 80% in Experiment 1; 86% in Experiment 2; and 87% in Experiment 3. The cutoff point of only 50% accuracy on comprehension questions is explained by two factors. First, comprehension questions were not primary-dependent variables but rather served to motivate participants to pay attention to the meaning of sentences. Second, there was very little evidence that participants' performance on the comprehension questions in all three experiments was related to the accuracy on the main task. In Experiment 1, participants with high accuracy on the comprehension task (above 80%) had an accuracy of 98% on the main task; participants with low accuracy on the comprehension task (below 80%) had an accuracy of 99% on the main task, χ^2 (5, N = 88) = 3.07, p = .689. In Experiment 2, participants with high accuracy on the comprehension task (above 80%) had an accuracy of 99% on the main task; participants with low accuracy on the comprehension task (below 80%) had an accuracy of 96% on the main task, χ^2 (5, N = 84) = 19.13, p = .002. In Experiment 3, participants with high accuracy on the comprehension task (above 80%) had an accuracy of 97% on the main task; participants with low accuracy on the comprehension task (below 80%) had an accuracy of 95% on the main task, χ^2 (10, N = 105) = 13.82, p = .181. Thus, in all three experiments participants with low accuracy on the comprehension task (below 80%) had an accuracy of more than 95% on the main task. Furthermore, only in one out of three experiments there was an indication that participants' performance on the comprehension questions was somehow related to the accuracy on the main task.
2. We thank an anonymous reviewer for this suggestion.
3. We thank an anonymous reviewer for this suggestion.

References

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The "blank screen paradigm". *Cognition*, *93*(2), B79–B87. https://doi.org/10.1016/j.cognition.2004.02.005

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. H., & 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

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*(5–6), 513–562. https://doi.org/10.1080/01690960344000026

Barsalou, L. W. (2015). Situated conceptualization: Theory and applications. In Y. Coello & M. H. Fischer (Eds.), *Foundations of embodied cognition: Perceptual and emotional embodiment* (pp. 11–37). East Sussex, UK: Psychology Press.

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, *31*(5), 733–764. https://doi.org/10.1080/03640210701530748

Borreggine, K. L., & Kaschak, M. P. (2006). The action-sentence compatibility effect: It's all in the timing. *Cognitive Science*, *30*(6), 1097–1112. https://doi.org/10.1207/s15516709cog0000_91

Braje, W. L., Legge, G. E., & Kersten, D. (2000). Invariant recognition of natural objects in the presence of shadows. *Perception*, *29*(4), 383–398. https://doi.org/10.1068/p3051

Castiello, U. (2001). Implicit processing of shadows. *Vision Research*, *41*(18), 2305–2309. https://doi.org/10.1016/S0042-6989(01)00141-9

Coppens, L. C., Gootjes, L., & Zwaan, R. A. (2012). Incidental picture exposure affects later reading: Evidence from the N400. *Brain and Language*, *122*(1), 64–69. https://doi.org/10.1016/j.bandl.2012.04.006

de Koning, B. B., Wassenburg, S. I., Bos, L. T., & Van der Schoot, M. (2017). Size does matter: Implied object size is mentally simulated during language comprehension. *Discourse Processes*, *54*(7), 493–503. https://doi.org/10.1080/0163853X.2015.1119604

Dee, H. M., & Santos, P. E. (2011). The perception and content of cast shadows: An interdisciplinary review. *Spatial Cognition & Computation*, *11*(3), 226–253. https://doi.org/10.1080/13875868.2011.565396

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

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Los Angeles, CA: Sage.

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*, *61*(6), 905–919. https://doi.org/10.1080/17470210701625550

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2013). *Multilevel and longitudinal modeling with IBM SPSS* (2nd ed.). New York, NY: Routledge.

Hoeben Mannaert, L. N., Dijkstra, K., & Zwaan, R. A. (2017). Is color an integral part of a rich mental simulation? *Memory & Cognition*, *45*(6), 974–982. https://doi.org/10.3758/s13421-017-0708-1

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*, 1–14. https://doi.org/10.3389/fpsyg.2016.00448

Horton, W. S., & Rapp, D. N. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review*, *10*(1), 104–110. https://doi.org/10.3758/BF03196473

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

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

Knoeferle, P., & Guerra, E. (2016). Visually situated language comprehension. *Language and Linguistics Compass*, *10*(2), 66–82. https://doi.org/10.1111/lnc3.12177

Laeng, B., & Sulutvedt, U. (2014). The eye pupil adjusts to imaginary light. *Psychological Science*, *25*(1), 188–197. https://doi.org/10.1177/0956797613503556

Leek, E. C., Davitt, L. I., & Cristino, F. (2015). Implicit encoding of extrinsic object properties in stored representations mediating recognition: Evidence from shadow-specific repetition priming. *Vision Research*, *108*, 49–55. https://doi.org/10.1016/j.visres.2015.01.011

Lynott, D., & Connell, L. (2010). Embodied conceptual combination. *Frontiers in Psychology*, *1*, 1–14. https://doi.org/10.3389/fpsyg.2010.00212

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology, Paris*, *102*(1–3), 59–70. https://doi.org/10.1016/j.jphysparis.2008.03.004

Mamassian, P., Knill, D. C., & Kersten, D. (1998). The perception of cast shadows. *Trends in Cognitive Sciences*, *2*(8), 288–295. https://doi.org/10.1016/s1364-6613(98)01204-2

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition*, *32*(8), 1389–1400. https://doi.org/10.3758/BF03206329

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

Montero-Melis, G., Isaksson, P., van Paridon, J., & Ostarek, M. (2019). Does using a foreign language reduce mental imagery? *Cognition*, *196*, 104134. https://doi.org/10.1016/j.cognition.2019.104134

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., & Huettig, F. (2019). Six challenges for embodiment research. *Current Directions in Psychological Science*, *28*(6), 593–599. https://doi.org/10.1177/0963721419866441

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

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 785–794. https://doi.org/10.1037/0278-7393.21.3.785

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *The European Journal of Neuroscience*, *21*(3), 793–797. https://doi.org/10.1111/j.1460-9568.2005.03900.x

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, *29*(6), 1045–1060. https://doi.org/10.1207/s15516709cog0000_29

Richardson, D., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition*, *102*(1), 129–138. https://doi.org/10.1016/j.cognition.2005.12.004

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Sato, K., Kanazawa, S., & Yamaguchi, M. K. (2017). Infants' perception of lightness changes related to cast shadows. *PLoS One*, *12*(3), e0173591. https://doi.org/10.1371/journal.pone.0173591

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

Schiffman, H. R. (1976). *Sensation and perception: An integrated approach* (3rd ed.). New York: John Wiley & Sons.

Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, *32*(2), 244–259. https://doi.org/10.3758/BF03196856

Speed, L. J., & Vigliocco, G. (2014). Eye movements reveal the dynamic simulation of speed in language. *Cognitive Science*, *38*(2), 367–382. https://doi.org/10.1111/cogs.12096
Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research Psychologische Forschung*, *65*(4), 235–241. https://doi.org/10.1007/s004260100059
Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156. https://doi.org/10.1111/1467-9280.00326
Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096–1101. https://doi.org/10.1038/nn1500
Williams, B. A. (1983). Another look at contrast in multiple schedules. *Journal of the Experimental Analysis of Behavior*, *39*(2), 345–384. https://doi.org/10.1901/jeab.1983.39-345
Wilson, N. L., & Gibbs, R. W. (2007). Real and imagined body movement primes metaphor comprehension. *Cognitive Science*, *31*(4), 721–731. https://doi.org/10.1080/15326900701399962
Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition*, *105* (1), 229–236. https://doi.org/10.1016/j.cognition.2006.09.003
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., 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

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Data S1. Results of multilevel models with different random effects structures for Experiments 1 to 3

Appendix A:

Samples of experimental and filler sentences from Experiments 1–3 (sentences in original Portuguese language are provided in parentheses)

Experimental sentences in Experiments 1 and 2

- You see how the sun is shining onto a backpack through the blinds/an open window (*Vê como o sol brilha na mochila através das persianas/da janela aberta*)
- You see how the sun is shining onto a teddy bear through the blinds/an open window (*Vê como o sol brilha no ursinho de peluche através das persianas/da janela aberta*)

Experimental sentences in Experiment 3

(identical except for "You see how" part)

- The sun is shining onto a backpack through the blinds/an open window (*O sol brilha na mochila através das persianas/da janela aberta*)
- The sun is shining onto a teddy bear through the blinds/an open window (*O sol brilha no ursinho de peluche através das persianas/da janela aberta*)

Filler sentences in Experiments 1 and 2

- You see how the sun is shining onto a pear through the clouds (*Vê como o sol brilha na pera através das nuvens*)
- You are opening a door with a key *(Tu estás a abrir a porta com uma chave)*

Filler sentences in Experiments 1 and 2 (identical except for "You see how" part in 12 of filler sentences)

- The sun is shining onto a pear through the clouds (*O sol brilha na pera através das nuvens*)
- You are opening a door with a key *(Tu estás a abrir a porta com uma chave)*