Psychonomic Bulletin & Review (2022) 29:1426–1439
https://doi.org/10.3758/s13423-022-02061-9

BRIEF REPORT

# Simulating background settings during spoken and written sentence comprehension

Oleksandr V. Horchak[1] · Margarida Vaz Garrido[1]

Accepted: 19 January 2022 / Published online: 7 February 2022

## Abstract

Previous findings from the sentence-picture verification task demonstrated that comprehenders simulate visual information about intrinsic attributes of described objects. Of interest is whether comprehenders may also simulate the setting in which an event takes place, such as, for example, the light information. To address this question, four experiments were conducted in which participants (total $N = 412$) either listened to (Experiment 1) or read (Experiment 3) sentences like "The sun is shining onto a bench" followed by a picture with the matching object (bench) and either the matching lighting condition of the scene (sunlit bench against the sunlit background) or the mismatching one (moonlit bench against the moonlit background). In both experiments, response times (RTs) were shorter when the lighting condition of the pictured scene matched the one implied in the sentence. However, no difference in RTs was observed when the processing of spoken sentences was interfered with visual noise (Experiment 2). Specifically, the results showed that visual interference disrupted incongruent visual content activated by listening to the sentences, as evidenced by faster responses on mismatching trials. Similarly, no difference in RTs was observed when the lighting condition of the pictured scene matched sentence context, but the target object presented for verification mismatched sentence context (Experiment 4). Thus, the locus of simulation effect is on the lighting representation of the target object rather than the lighting representation of the background. These findings support embodied and situated accounts of cognition, suggesting that comprehenders do not simulate objects independently of background settings.

**Keywords** Language comprehension · Visual simulation · Embodied cognition · Background settings · Light

## Introduction

Does language comprehension rely on visual simulation as suggested by perceptual symbol theories (Barsalou, 1999, 2008)? Much behavioral research has sought to answer this question using a sentence-picture verification paradigm (see Horchak et al., 2014, for a review). As one example, Zwaan et al. (2002) observed faster responses when the pictured object shape was compatible with the shape implied by the preceding sentence. As a different example, Winter and Bergen (2012) showed that verifying pictures depicting smaller objects was faster when reading sentences about distant objects than about nearby objects, and the reverse for the time to verify pictures depicting larger objects. The result that response times (RTs) are shorter whenever the pictured object matches the state implied by the sentence was taken as support for the hypothesis that people rely on visual simulation during the task.

Nonetheless, the above evidence could be interpreted differently. For example, comprehenders might not simulate an object as being in a specific state before picture verification. Instead, they might simply find it easier to incorporate the pictured version of the object when it matches sentence content (Masson, 2015). This explanation fits with the mechanism of backward semantic priming, according to which processing of picture stimuli should be supported by recruitment of the previously processed sentence stimuli (e.g., Neely et al., 1989). One of the most common mechanisms underlying semantic priming is spreading activation (Collins & Loftus, 1975), which suggests that there are strong links between the representations of related words in semantic memory. For example, reading a word such as "table" should activate the corresponding node in semantic

✉ Oleksandr V. Horchak
Oleksandr.Horchak@iscte-iul.pt

[1] Iscte-Instituto Universitário de Lisboa, Cis-Iscte, Av. das Forças Armadas, 1649-026 Lisbon, Portugal

memory that spreads to the words with similar meaning via the nearby nodes. Consequently, RTs for the word "stool" should be faster than RTs for the word "squirrel."

Recently, a more nuanced picture of the functional role of simulation during word processing has emerged with the use of visual noise. By using this technique, the assumed simulation is interfered with rapidly flashing visual masks that selectively activate the visual cortex (Yuval-Greenberg & Heeger, 2013), and the impact of this interference on the task is assessed. For example, Edmiston and Lupyan (2017) asked participants to listen to a word followed by the presentation of two pictured objects, one of which was oriented upright and the other was oriented upside down. Seventy-five percent of the time, the pictured objects matched the word (e.g., verifying pictures of two dogs after hearing "dog"), but 25% of the time, the pictured objects mismatched the word (e.g., verifying pictures of two cats after hearing "dog"). On 50% of all trials, participants saw visual noise in the form of colorful rectangles with colors, sizes, and positions alternating at a rate of around 60 Hz. Participants' task was to press the button corresponding to the side that displayed the image in upright position. The results showed that RTs for matching stimuli were approximately the same for trials with and without visual interference. However, RTs for mismatching stimuli were reduced for trials with (vs. without) visual interference. Edmiston and Lupyan (2017) concluded that visual noise disrupted incongruent visual content while listening to the word. Furthermore, in the same study the researchers measured the effect of visual interference on comprehenders' ability to answer questions about objects' properties. The results showed that visual interference reduced the accuracy in answering visual questions (e.g., color) but not non-visual questions (e.g., tactile feelings). Thus, Experiment 2 showed that visual interference affects only visual knowledge (see also Ostarek & Huettig, 2017, for further evidence).

Whereas the case for visual simulation is strong regarding word processing, the case for the involvement of visual processes during sentence processing is weaker. For example, Ostarek et al. (2019) investigated which processes contribute to the retrieval of shape information in a sentence-picture verification task by using the materials from the original Zwaan et al.'s (2002) study. They hypothesized that if faster RTs are explained by visual simulation, then visual interference occurring before the presentation of the target image should reduce the effect of the sentence on subsequent image recognition. The researchers found no evidence that disrupting visual processes interfered with visual simulation. This is the case because RTs were faster for shape-matching trials in both "blank screen" and "visual interference" conditions.

The above findings prompt further questions regarding the situations when visual processes are functionally involved during sentence processing. One possibility is that comprehenders need to rely on visual simulation when a sentence describes a more complex scene that includes the surrounding environment and any relevant objects. According to the simulation hypothesis (Barsalou, 2003, 2016), when attention focuses on any kind of object during real-life experience, then a simulator that develops for this object (i.e., a multimodal representation of the category) should include not only the object-specific information but also the setting where the event takes place. This view is supported by some empirical evidence. As one example, Yaxley and Zwaan (2007) demonstrated shorter RTs when the visual resolution of the depicted object matched the degree of object visibility implied by the sentence. As a different example, Horchak and Garrido (2020) found shorter RTs for pictures depicting objects with an alternating light pattern when preceded by sentences mentioning blinds. A limitation is that picture verification in these studies occurs only after sentence processing, thus making an alternative interpretation based on retroactive mechanisms in priming a viable possibility. However, demonstrating that interfering with visual processing leads to a different pattern of results (e.g., no advantage for matching trials) would provide a stronger argument for the view that comprehenders visually simulate the situation implied by the sentence. The work reported in this article was designed to provide such evidence.

In the present research, we addressed the importance of background information regarding the simulation of light. To do so, we manipulated light information by asking participants to listen to (Experiment 1) or read (Experiment 3) sentences such as "The sun/the moon is shining onto a bench" followed by a picture with the matching object and either the matching lighting condition of the scene or the mismatching one. To probe for the involvement of visual processes, we used low-level visual noise during the presentation of spoken sentences (Experiment 2) and showed pictures with mismatching objects but matching lighting conditions after the presentation of written sentences (Experiment 4). If participants simulate light information, we expect to see an interaction such that responding is faster when both object and setting information from the picture match sentence content. At the same time, if responding requires activation of visual representations, then we would not expect to observe faster responses for matching trials when the assumed simulation is disrupted by visual interference, as well as when picture stimuli are compatible on only one dimension (e.g., compatible background but incompatible target object) with sentence content.

# Experiment 1

## Method

### Power analysis

We performed a simulation-based power analysis to calculate the number of participants needed to detect the critical interaction between sentence type and picture type. This approach requires running an experiment many times and calculating the proportion of statistically significant results. Specifically, we used the "mixedpower" package of Kumle et al. (2018) on the data (Experiment 7) published by Horchak and Garrido (2021), where the main finding was the significant interaction between sentences and pictures such that the state of the object implied by the sentence influenced verification responses. Our power estimation followed the recommendations described by Kumle et al. (2021). Specifically, it consisted of the following steps. As a starting point, we fitted the linear mixed-effects model on the data, where sentence type, picture type, and their interaction were fixed effects; and participants and items were random effects. Then, we estimated a power of 80% over a range of different sample sizes (50, 70, 90, 100, 120); defined a $t$-value of 2 as our threshold of significance (Baayen et al., 2008); and "instructed" the model to run 1,000 repetitions in the simulation process, which is the default value in all functions of the "mixedpower" package. Although Horchak and Garrido (2021) observed a robust interaction effect, relying on the exact data-based estimations is undesirable due to other non-methodological differences between two studies (e.g., different research idea, materials, etc.). Therefore, to account for uncertainty in the data and reduce the unknown risk of anticonservativity, we determined our smallest effect size of interest (SESOI) by reducing all beta coefficients for fixed effects by 20%. This approach is similar to that described by Kumle et al. (2021), where SESOI was determined by reducing all beta coefficients by 15%. Simulation results suggested that we would need at least 90 participants for each experiment to detect the "interaction" effect between sentences and pictures if it existed.

### Norming study

As we were interested in testing whether comprehenders situate the category in background settings, it was important that the targets depicted in the pictures were familiar and grounded in naturalistic contexts. To this end, we selected the names of objects and animals based on their high imageability scores ($M > 6.00$ on a 7-point scale) from the Glasgow Norms ratings (Scott et al., 2019). Then, we created a list of 11 light sources (e.g., sunlight, fireworks, torch, stars, etc.) and asked 99 participants (82 females, $M_{\text{age}} =$ 23.9 years) to identify perceptual contexts within which observing objects or animals most often occurs.[1] Notably, we did not include sources of light that have more than one dominant color associated with them. For example, we did not include streetlights as they may imply both warm and cold colors, and it is not possible to predict what kind of streetlights participants typically see in their lives. Each light source should receive a "frequency" rating above 4 on the 7-point scale (1 = Not frequent at all; 7 = Very frequent) to be used in the experiments. The data showed that sunlight and moonlight were the only sources of light that met this requirement ($M_{\text{sun}} = 6.54$; $M_{\text{moon}} = 5.26$). Finally, it was also necessary to ensure that the findings were not confounded by the degree to which a background setting was associated with a specific color (Tanaka & Presnell, 1999). To this end, we presented 106 new participants (91 females, $M_{\text{age}} = 23.1$ years) with all experimental sentences and pictures (one sentence-picture pair at a time) and asked them to evaluate the quality of the pictures regarding their match with sentence content[2] on a 7-point scale (1 = Very low; 7 = Very high). There was no effect of background setting on quality ratings ($M_{\text{moon}} = 5.33$; $M_{\text{sun}} = 5.30$, $t\ (105) = 0.51$, $p = .611$, $d = .050$).

### Participants

Ninety-eight undergraduate university students (all were native speakers of Portuguese) took part in Experiment 1 in exchange for course credit. Because of the coronavirus pandemic 2019 (COVID-19), students in this and all subsequent experiments signed up for a study online through the Sona Systems cloud-based software. The responses of nine participants were eliminated due to low accuracy (<80%). Thus, the results of Experiment 1 are based on the data from 89 participants ($M_{\text{age}} = 20.86$ years, $SD_{\text{age}} = 5.37$), of whom 74 were females.

### Materials

We created 24 experimental sentence pairs and 48 filler sentences. All experimental sentences were of the form *"The*

[1] The instructions were as follows: In everyday life, we observe the world in different lighting conditions. Based on your experience, how often do we observe objects and animals in the following lighting conditions? Please indicate your response on a scale from 1 (Not often at all) to 7 (Very often).

[2] The instructions were as follows: You will be presented with different sentence-picture pairs (one sentence-picture pair at a time). Your task is to evaluate the quality of the picture in terms of how well it matches the situation described in the sentence. Please indicate your response on a scale from 1 (Very low quality) to 7 (Very high quality).

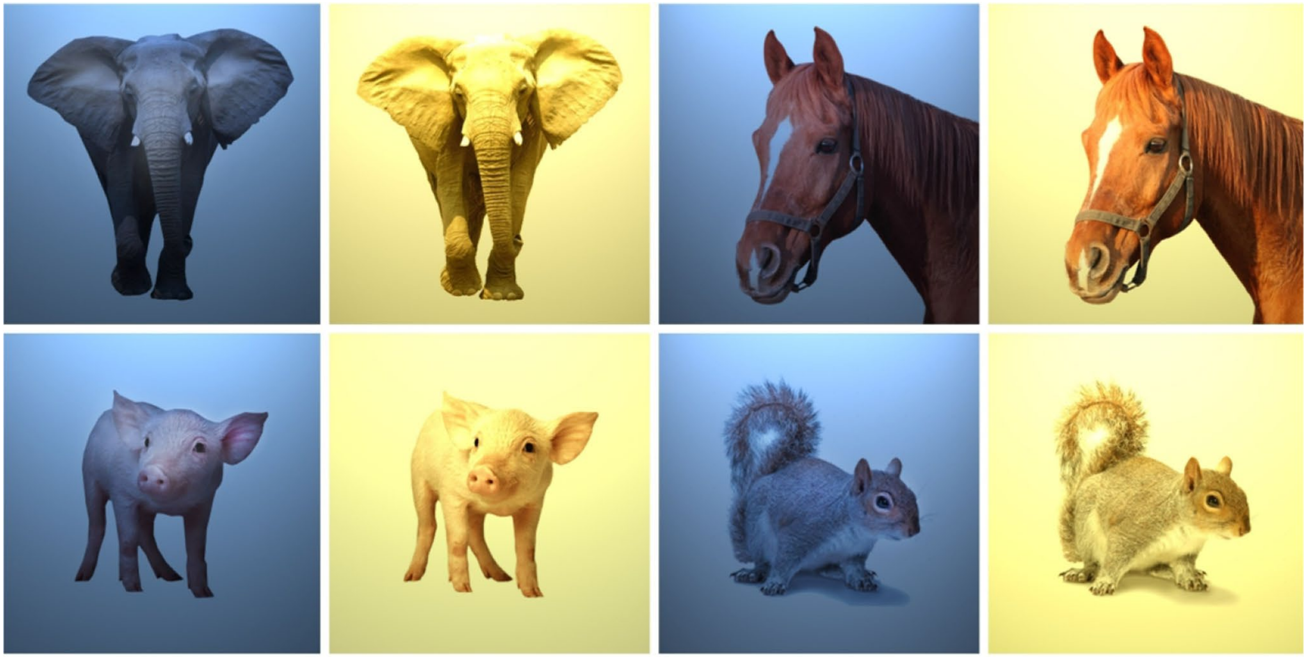

**Fig. 1** Examples of target objects situated in sunlit and moonlit background settings

*sun/the moon is shining onto object X."* Thus, we varied the background setting in which the object is situated. For example, the sentence *"The sun is shining onto a bench"* implies that a bench resides in a warm light setting, whereas the sentence *"The moon is shining onto a bench"* implies a cold light setting. All experimental sentences were followed by a pictured object mentioned in a sentence and required a "yes" response. Twenty-four of 48 filler sentences were the same as experimental sentences, except they were followed by a pictured object not mentioned in a sentence and required a "no" response. The remaining 24 sentences included other sources of light (e.g., torch, stars, fireworks, etc.) and required equal numbers of "yes" and "no" responses. Overall, there were 36 trials requiring a "yes" response (24 experimental and 12 filler items) and 36 trials requiring a "no" response (all filler items). All sentences were presented in European Portuguese. They were recorded by a male native speaker at a normal reading rate and were approximately 2,500 ms in duration. Finally, to motivate participants to listen to sentences attentively, we also created 24 comprehension questions[3] that appeared after half of all filler trials (e.g., The light from the stars was shining onto a bench?).

We created same-sized images of scenes (385 × 385 pixels) to go with each sentence: 24 experimental picture pairs and 48 filler pictures. Both members of each experimental pair depicted the same object except for the background setting (sunlit vs. moonlit) in which the object is situated. The other 48 pictures were fillers, with half of the pictures depicting a sunlit object against a sunlit background and the other half depicting a moonlit object against a moonlit background. Sunlit and moonlit backgrounds (see Fig. 1) were applied using Adobe Photoshop (Concepcion, 2019).

## Design

In this and all subsequent experiments, there were four lists of stimuli, with each experimental sentence-picture pair appearing in only one of the following conditions per list: sun sentence-sunlit picture background; sun sentence-moonlit picture background; moon sentence-sunlit picture background; and moon sentence-moonlit picture background. Each participant verified items that appeared in all four conditions, but each item appeared in only one condition per list, and each participant was randomly assigned to only one list. As the counterbalanced list was of little theoretical interest to us, it was not included as a factor in the statistical modeling. Thus, the present research employed a 2 (sentence: sun vs. moon) × 2 (picture: sunlit background vs. moonlit background) within-participants design.

## Procedure

Participants were instructed to perform a task requiring them to listen to a sentence through the headphones and then

[3] These questions were not primary dependent variables to us. However, the mean accuracy of all participants was always above 50%.

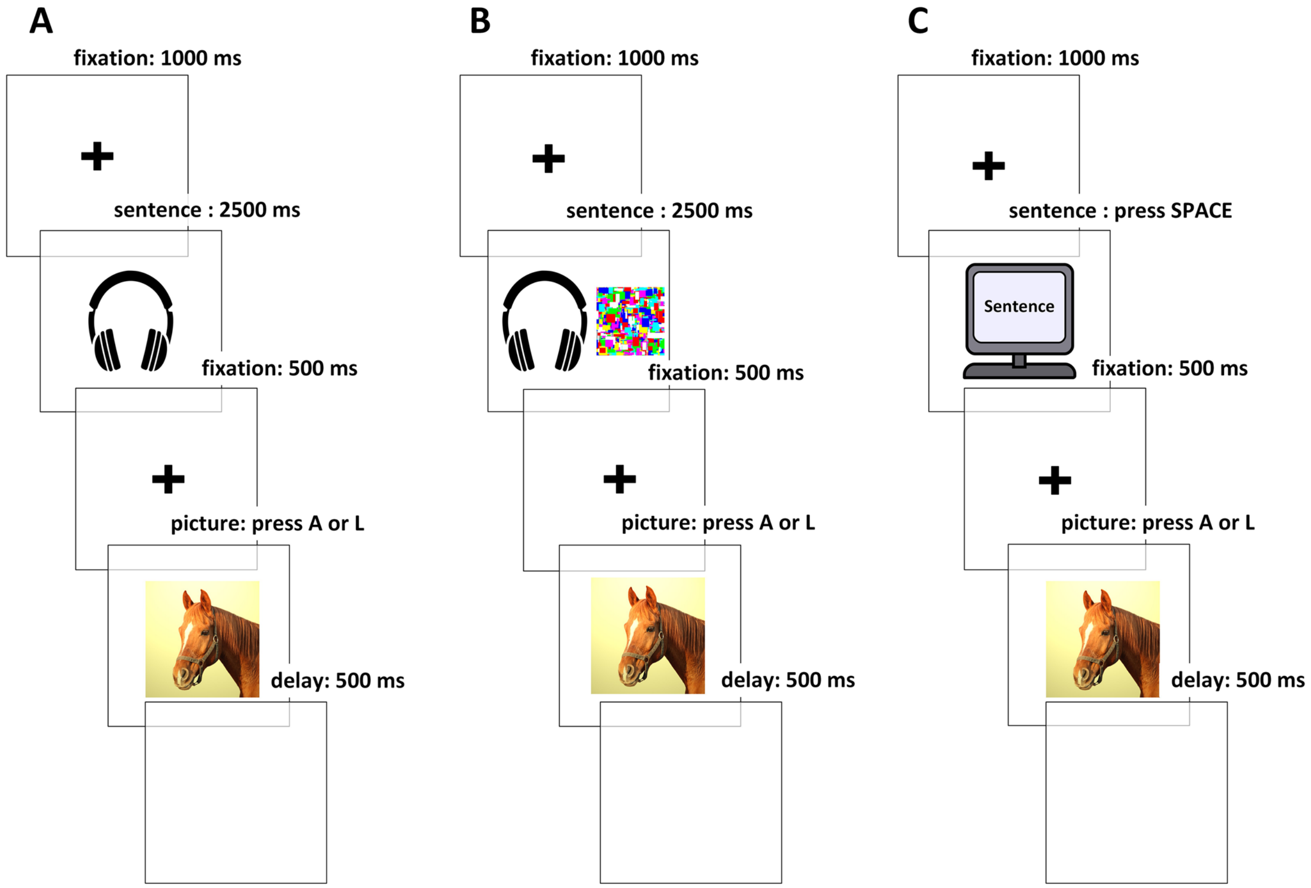

**Fig. 2** Representation of the trial sequence in Experiments 1–4. *Note.* **a** A sample trial from Experiment 1 with auditory sentences. **b** A sample trial from Experiment 2, which was identical to Experiment 1, except that auditory sentences were accompanied by visual noise. **c** A sample trial from Experiments 3 and 4 with written sentences

decide whether the subsequently presented pictured object had been mentioned in the sentence. In addition, instructions warned participants about the need to listen to the sentences attentively as their comprehension would be tested.

As demonstrated in Fig. 2a, the experiment began with eight practice trials, where participants received visual feedback on the accuracy of their responses. On each trial (both practice and main), participants heard a sentence followed by a picture depicting a target that either resided in a warm light setting or a cold light setting. Each trial started with a fixation cross displayed for 1,000 milliseconds (ms) in the center of the screen, after which an auditory sentence was played (approximately 2500 ms in duration). After the offset of the sentence, a fixation cross appeared for 500 ms, immediately followed by a pictured object. Then, participants indicated whether the pictured target was mentioned in the preceding sentence. Specifically, they pressed the button “L” to indicate a “yes” response and the button “A” to indicate a “no” response. Finally, there were 24 comprehension questions presented after half of the filler pictures, with “yes” and “no” responses being required an equal number of times.

Stimuli delivery was controlled by a web-based service PsyToolkit (Stoet, 2010, 2017). The advantage of this service is that all stimuli are loaded into the participants’ computers before the experiment starts, thus ensuring that there are no delays due to internet connection. Kim et al. (2019) found that the data collected using Psytookit are comparable to the data collected using E-Prime 3.0 in a complex psycholinguistic task. In the present research, participants could only start the experiment if their web browser supported a full-screen mode. Furthermore, they could only access the study via a desktop computer or a laptop (i.e., smartphones and tablets were not permitted).

## Data treatment

In line with previous similar studies (e.g., Connell, 2007; de Koning et al., 2017; Zwaan & Pecher, 2012), and in all four experiments, we first removed all filler items and the data from participants with an overall accuracy of less than 80% on experimental trials. For RT analyses, we omitted all incorrect responses and then discarded responses faster than 300 ms and slower than 3,000 ms, as well as responses with RTs 2.5 SDs higher from the relevant condition's mean. Finally, we checked response times (RTs) for normality and found that RTs in this and all subsequent experiments were positively skewed, hence violating the assumption of normally distributed variables. Thus, we applied logarithmic (log10) transformation[4,5] to get normal distributions.

## Data analysis

All statistical analyses in Experiments 1–4 were performed within the R programming environment version 4.0.5 (R Core Team, 2020) and several R packages.[6] Accuracy scores and RTs were analyzed with logistic and linear mixed-effects regression models,[7] respectively. To reduce the unknown risk of anticonservativity, we fitted the "maximal" random-effects structure justified by the experimental design (Barr et al., 2013). The full model included sentence type, picture type, and their interaction as fixed effects; by-participant and by-item random intercepts, as well as by-participants slopes for sentence type, picture type, and the interaction term as random effects. In the case of non-convergence of the "maximal" model, we first "de-correlated" the intercept and slope, and if it did not work, we removed terms required to allow a successful convergence. Fixed effects predictors were sum-coded (1, -1) to facilitate the interpretation of main effects in the presence of interactions. In the presence of a significant interaction, we used dummy coding of the picture condition factor to obtain the simple effects of sentence condition on "sunlit" and "moonlit" trials.

## Results and discussion

Participants' overall accuracy in all experiments was always higher than 97%.[8] No significant effects were found for accuracy ($z$ <2). Regarding RTs, there were no main effects of sentence and picture type in any of the four experiments.[9]

[4] We also ran the analyses on raw RTs and found that the same results were significant both in the "transformed" and the "untransformed" analysis. In Experiment 1, there was a significant interaction between sentences and pictures (*estimate* = −17.392, *SE* = 4.666, *t* = −3.728, *p* < .001), with faster RTs for sunlit pictures when preceded by "sun" sentences (*estimate* = 20.617, *SE* = 6.883, *t* = 2.995, *p* = .003) and moonlit pictures when preceded by "moon" sentences (*estimate* = −14.167, *SE* = 6.879, *t* = −2.059, *p* = .040). In Experiment 2, the interaction between sentences and pictures was not significant (*estimate* = −0.539, *SE* = 4.320, *t* = −0.125, *p* = .901). In Experiment 3, there was a significant interaction between sentences and pictures (*estimate* = −16.464, *SE* = 4.811, *t* = −3.422, *p* < .001), with faster RTs for moonlit pictures when preceded by "moon" sentences (*estimate* = −16.228, *SE* = 6.939, *t* = −2.339, *p* = .020) and sunlit pictures when preceded by "sun" sentences (*estimate* = 16.700, *SE* = 6.940, *t* = 2.407, *p* = .017). In Experiment 4, the interaction between sentences and pictures was not significant (*estimate* = −4.082, *SE* = 4.668, *t* = −0.874, *p* = .382).

[5] As kindly suggested by one of the reviewers, we also performed a Box-Cox transformation to make sure that the observed RT results did not depend on doing log transformation. The purpose of Box-Cox transformation is to identify an appropriate exponent (Lambda) to use to transform data into a "normal shape." In all four experiments the best values for Lambda were in the range from -0.52 to -0.58 (confidence intervals did not include whole numbers like 0 and 1), and thus we chose a Lambda value of -0.5 as the power to which all data should be raised. The results showed that the results using "Lambda" RTs were similar to those using log RTs and raw RTs. In Experiment 1, there was a significant interaction between sentences and pictures (*estimate* = 0.0004, *SE* = 0.0001, *t* = 3.350, *p* < .001), with faster RTs for sunlit pictures when preceded by "sun" sentences (*estimate* = -0.0005, *SE* = 0.0002, *t* = -2.808, *p* = .005) and moonlit pictures when preceded by "moon" sentences (*estimate* = 0.0003, *SE* = 0.0002, *t* = 1.932, *p* = .054). In Experiment 2, the interaction between sentences and pictures was not significant (*estimate* = 0.0001, *SE* = 0.0001, *t* = 0.456, *p* = .649). In Experiment 3, there was a significant interaction between sentences and pictures (*estimate* = 0.0004, *SE* = 0.0001, *t* = 2.656, *p* = .009), with faster RTs for sunlit pictures when preceded by "sun" sentences (*estimate* = -0.0004, *SE* = 0.0002, *t* = −1.984, *p* = .050) and moonlit pictures when preceded by "moon" sentences (*estimate* = 0.0003, *SE* = 0.0002, *t* = 2.025, *p* = .044). In Experiment 4, the interaction between sentences and pictures was not significant (*estimate* = 0.0001, *SE* = 0.0001, *t* = 1.213, *p* = .225).

[6] The "tidyverse" package (Wickham et al., 2019) was used for data wrangling; and the "lme4" package (Bates et al., 2015) and "lmerTest" package (Kuznetsova et al., 2017) were used for main statistical analyses.

[7] Generalized linear mixed model (family binomial) was used to analyze accuracy with the formula: Accuracy ~ sentence * picture + (1 + sentence * picture | ppt) + (1 | item). Linear mixed model (fit by REML) was used to analyze RTs with the formula: log. RT ~ sentence * picture + (1 + sentence * picture | ppt) + (1 | item).

[8] In the present research, and consistent with previous similar studies (e.g., Connell, 2007; de Koning et al., 2017; Zwaan & Pecher, 2012), we excluded participants if their accuracy threshold was lower than 80%. At the request of a reviewer, we also ran the analyses using all the data (we only excluded two participants with accuracy < 50%) to check if the critical interaction between sentences and pictures is still observed. As for accuracy, the interaction was not significant for Experiment 2 (*estimate* = −0.194, *SE* = 0.496, *z* = −0.391, *p* = .696), Experiment 3 (*estimate* = −0.223, *SE* = 0.291, *z* = −0.768, *p* = .443), and Experiment 4 (*estimate* = −0.265, *SE* = 0.446, *z* = −0.594, *p* = .552). However, it was significant for Experiment 1 (*estimate* = −1.735, *SE* = 0.719, *z* = −2.414, *p* = .016), reflecting the fact that participants were more accurate in verifying a sunlit picture after reading a "sun" sentence (*M* = 0.98; *SD* = 0.15) than a "moon" sentence (*M* = 0.94; *SD* = 0.25); and a moonlit picture after reading a "moon" (*M* = 0.97; *SD* = 0.16) sentence than a "sun" (*M* = 0.93; *SD* = 0.25) sentence. Regarding RTs, the results for the critical interaction were similar. Specifically, there was an interaction between sentences and pictures in Experiment 1 (*estimate* = −0.009, *SE* = 0.002, *t* = −3.113, *p* = .003) and Experiment 3 (*estimate* = – -0.008, *SE* = 0.003, *t* = – -2.625, *p* = .010). However, no interaction was observed in Experiment 2 (*estimate* = – 0.002, *SE* = 0.003, *t* = −0.610, *p* = .542) and Experiment 4 (*estimate* = – 0.003, *SE* = 0.002, *t* = −1.241, *p* = .215).

[9] In Experiment 1, there were no main effects of sentence type (*estimate* = 0.002, *SE* = 0.003, *t* = 0.640, *p* = .524) and picture type (*estimate* = 0.002, *SE* = 0.003, *t* = 0.581, *p* = .562). In Experiment 2, there were no main effects of sentence type (*estimate* = 0.001, *SE* = 0.003, *t* = 0.367, *p* = .714) and picture type (*estimate* = −0.001, *SE*

Thus, the results section of each experiment is focused on the analysis of the critical interaction of interest between sentences and pictures for RT data (see Appendix 1 for more information about accuracy and RT data).

As shown in Fig. 2, there was a significant interaction between sentences and pictures (*estimate* = −0.010, *SE* = 0.003, *t* = −3.347, *p* = .001) in Experiment 1. Follow-up analyses showed that moonlit pictures were responded to faster when preceded by a "moon" sentence (*estimate* = −0.008, *SE* = 0.004, *t* = −2.013, *p* = .045), and sunlit pictures were responded to faster when preceded by a "sun" sentence (*estimate* = 0.011, *SE* = 0.004, *t* = 2.862, *p* = .005).

One could argue that these data merely point to the informational content activated during sentence processing, but are silent on the specific mental mechanisms underlying such activation. Therefore, in Experiment 2, we used visual interference during the presentation of the spoken sentences to investigate whether there is a reduction or an elimination of the RT difference between matching and mismatching conditions when simulation is prevented by visual noise. If the difference in RTs from Experiment 1 is due to response facilitation in the matching condition, RTs for the matching condition should increase. This is the case because visual interference should disrupt congruent visual content activated by listening to a sentence, content that otherwise facilitates verifying the picture. If, however, the difference in RTs from Experiment 1 is due to response inhibition in the mismatching condition, RTs for the mismatching condition should decrease. This is the case because visual noise should disrupt incongruent visual content activated by listening to the sentence, content that otherwise hinders verifying the picture.

# Experiment 2

## Method

### Participants

We recruited 106 native-speaking university students via Sona Systems software in exchange for course credit. Responses of eight participants with accuracy <80% were eliminated. Thus, main analyses were run on the data from 98 participants ($M_{\text{age}}$ = 21.08 years, $SD_{\text{age}}$ = 5.48), of whom 74 were females.

### Materials

Materials of Experiment 2 were identical to Experiment 1, except that 40 Mondrian-type masks were created by superimposing many rectangles of different sizes and colors. The colors of the rectangles were similar to those used in Edmiston and Lupyan (2017).

### Procedure

The procedure of Experiment 2 was identical to Experiment 1, except that all experimental sentences and 12 filler sentences (that is, half of all trials) were accompanied by visual noise. This noise consisted of 40 masks that were alternating at a rate of 60 Hz (see Fig. 3).

## Results and discussion

The interaction between sentences and picture was not significant when sentences were accompanied by visual noise (*estimate* = −0.001, *SE* = 0.003, *t* = −0.374, *p* = .709). The pattern of observed RTs (see Fig. 2) demonstrates that this occurred primarily due to faster responses on mismatching trials. This suggests that visual interference disrupted incongruent visual content activated by listening to the sentences. The follow-up analysis over RT data from both experiments[10] showed that there was no main effect of experiment (*estimate* = 0.003, SE = 0.007, *t* = 0.378, *p* = .706). However, there was a significant three-way interaction between sentences, pictures, and experiments (*estimate* = −0.004, *SE* = 0.002, *t* = −2.112, *p* = .036). Thus, visual interference disrupted, even if partially, visual representations.

# Experiment 3

Experiments 1 and 2 provide evidence for the simulation of background information during spoken sentence comprehension. Experiments 3 and 4 were designed to provide the same evidence for written sentence comprehension.

Footnote 9 (continued)

= 0.003, *t* = −0.436, *p* = .663). In Experiment 3, there were no main effects of sentence type (*estimate* = 0.000, *SE* = 0.003, *t* = 0.112, *p* = .911) and picture type (*estimate* = −0.003, *SE* = 0.003, *t* = −1.068, *p* = .287). Finally, in Experiment 4, there again were no main effects of sentence type (*estimate* = 0.001 *SE* = 0.002, *t* = 0.284, *p* = .776) and picture type (*estimate* = −0.001, *SE* = 0.003, *t* = −0.330, *p* = .742).

[10] For this analysis, we used the same linear mixed-effect model as before, except that we added the "experiment" factor to the model. Thus, the formula was: *log.RT ~ sentence * picture * experiment + (1 + sentence * picture | ppt) + (1 | item).*

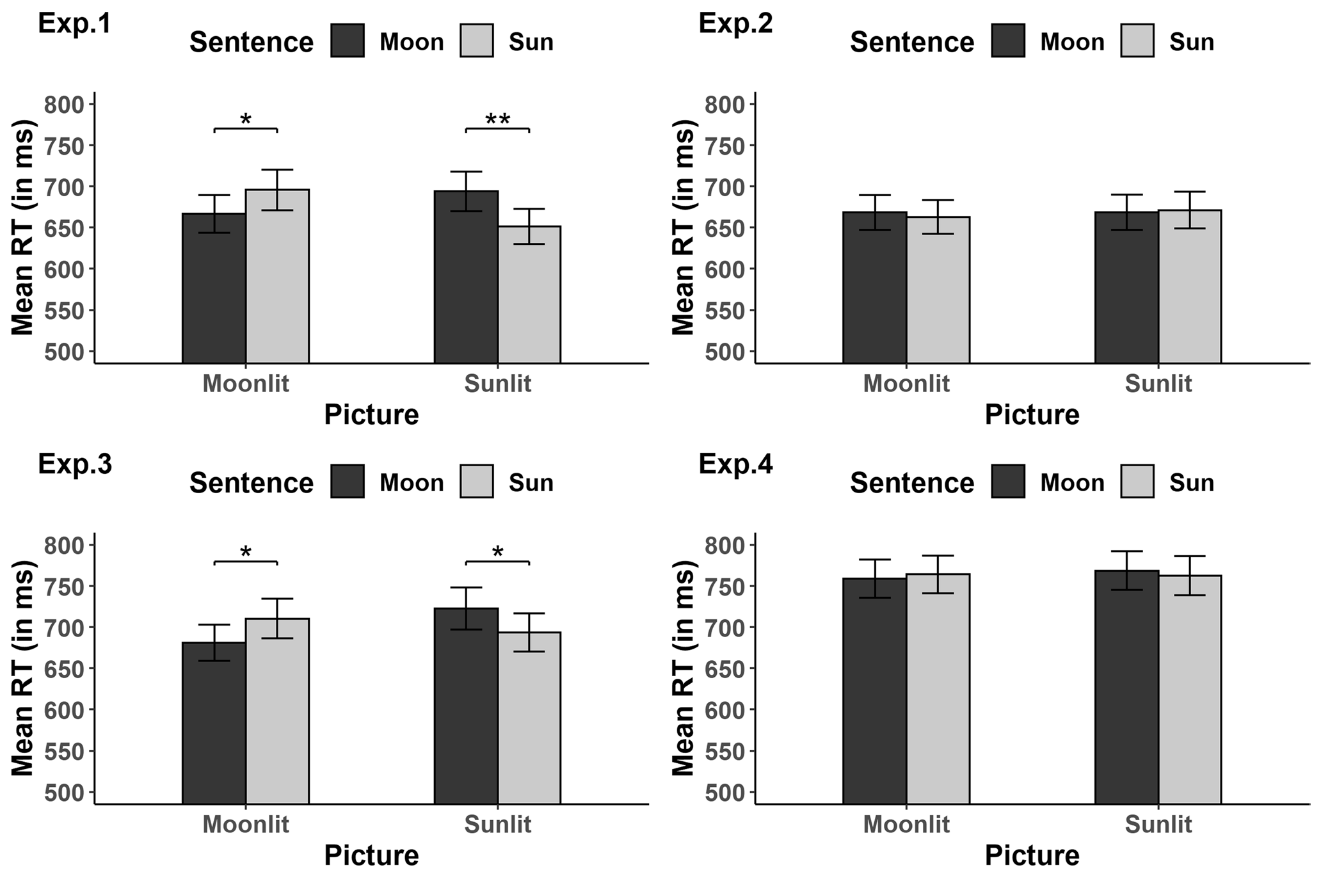

**Fig. 3** Mean response times in milliseconds with 95% confidence intervals (Experiments 1–4). *Note.* (**Exp.1**) Results of Experiment 1, in which participants listened to the sentences and then verified pictures with the matching object and either the matching lighting condition of the scene or the mismatching one. (**Exp.2**) Results of Experiment 2, which was identical to Experiment 1, except that sentences were accompanied by visual noise. (**Exp.3**) Results of Experiment 3, which was identical to Experiment 1, except that participants read the sentences presented in the middle of the screen. (**Exp.4**) Results of Experiment 4, which was nearly identical to Experiment 3, except that participants verified pictures with the mismatching target object. ***p* < .01. **p* < .05

## Method

### Participants

We recruited 100 native-speaking university students via Sona Systems software in exchange for course credit. Responses of ten participants with accuracy <80% were discarded. Thus, the analyses were run on the data from 90 participants ($M_{\text{age}}$ = 20.88 years, $SD_{\text{age}}$ = 5.07), of whom 71 were females.

### Materials

Materials of Experiment 3 were identical to those used in previous experiments, except that participants were instructed to read the sentences presented in the middle of the screen.

### Procedure

The procedure of Experiment 3 was nearly identical to Experiment 1. Specifically, each trial started with a fixation cross in the middle of a screen for 1,000 ms, followed by a sentence in the middle of the screen. The sentence remained on the screen until participants pressed the spacebar to indicate that they had read and understood the sentence. After a spacebar press, the sentence was replaced by a fixation cross for 500 ms, immediately followed by a pictured object. The task was the same as in the previous two experiments.

## Results and discussion

As in Experiment 1, there was a significant interaction between sentences and pictures (*estimate* = –0.009, *SE* = 0.003, *t* = –2.821, *p* = .006). As demonstrated in Fig. 2, follow-up analyses showed that moonlit pictures were responded to more quickly when preceded by a "moon"

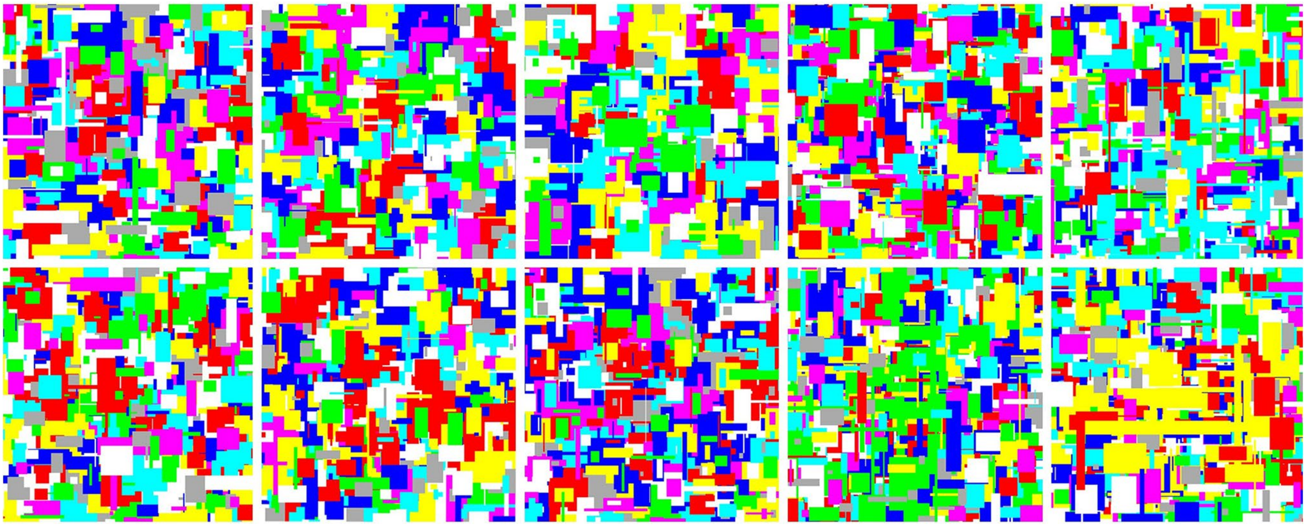

**Fig. 4** Examples of visual masks used in Experiment 2

sentence (*estimate* = –0.009, *SE* = 0.004, *t* = –2.191, *p* = .030), and sunlit pictures were responded to more quickly when preceded by a "sun" sentence (*estimate* = 0.009, *SE* = 0.004, *t* = 2.050, *p* = .043). Thus, these results replicate those from Experiment 1. However, they do not reveal which processes enable the retrieval of background information. Experiment 4 was designed to address this issue.

## Experiment 4

The aim of Experiment 4 was to test the involvement of visual processes in written sentence comprehension. We did not use visual noise like in Experiment 2 because of the concern that participants could develop a strategy to selectively focus on the part of the screen where sentences are presented and thus ignore visual interference. Instead, we used pictures in which only background information matched sentence context.

According to the simulation hypothesis, comprehenders integrate information from an object and its background (Barsalou, 2016). If this is the case, then it is not just the lighting representation of the background that should play a role in the speed of picture verification, but also the target object superimposed with the specific light source. Thus, the prediction is that in a sentence like "The sun is shining onto a *bench,*" a comprehender should form a visual representation of a described scene that involves a sunlit bench. If the picture presented for verification depicts, for example, a *horse*, then RTs should be approximately the same to the sunlit horse and the moonlit horse (both requiring a "no" response) since no simulation of light on a horse is required by the sentence. This prediction is supported by two lines of evidence and task requirements. First, empirical evidence suggests that target entities attract a greater level of attention relative to background information and hence contribute substantially to the interpretation of the scene early in processing (e.g., Biederman, 1972; Potter, 1975). Second, research shows that much visual information is required to process scenes with a low semantic similarity between objects and backgrounds (which is true for the present research, see Fig. 4) compared to scenes with high semantic similarity (e.g., Davenport & Potter, 2004). Finally, our task required participants to verify whether the object from the picture (and not the background) was mentioned in the preceding sentence.

By contrast, the amodal hypothesis suggests that sentence processing activates lists of category features in a semantic network to which the depicted picture is then compared. Specifically, a classical semantic priming account would predict facilitation in responding to a "sunlit" picture due to the semantically related word "sun," regardless of the target object being displayed. However, if the task suggests that the correct response is "no," then a sunny background becomes a distractor that needs to be suppressed. Hence, a comprehender needs extra time to overcome the distractor and respond to the pictured target correctly (see Neill & Valdes, 1992, for the mechanism of negative priming). In line with these theories, after the sentence mentioning "sun," RTs to a non-present sunlit object should be longer than to a non-present moonlit object.

## Method

### Participants

We recruited 108 native-speaking university students via Sona Systems software in exchange for course credit. Responses of six participants with accuracy <80% were discarded. Thus, the analyses were run on the data from 102 participants ($M_{\text{age}}$ = 21.06 years, $SD_{\text{age}}$ = 5.67), of whom 83 were females.

### Materials

Picture materials of Experiment 4 were the same as in all other experiments. However, experimental sentence stimuli mentioned the object that was not depicted in the subsequently presented picture (e.g., reading a sentence about how the sun is shining onto a box and then verifying a picture with a sunlit bench). That is, in the present experiment, a correct response for experimental trials was "no" ("A" button press). Furthermore, to allow for an even number of trials requiring "yes" and "no" responses, 24 filler "sun" and "moon" sentences now mentioned an object that matched the one from the picture (thus requiring a "yes" response).

### Procedure

The procedure was identical to Experiment 3.

## Results and discussion

Central to our prediction, the interaction between sentences and pictures was not significant (*estimate* = −0.003, *SE* = 0.002, $t$ = −1.132, $p$ = .258) when the target object from the picture mismatched that mentioned in the sentence. To get a better understanding of the differences among results, two follow-up analyses over RT data from the other experiments were performed.[11] The first analysis comparing the results from Experiments 1 and 4 showed a nearly significant three-way interaction between sentences, pictures, and experiments (*estimate* = −0.003, *SE* = 0.002, $t$ = −1.964, $p$ = .051). The second analysis comparing the results from Experiments 3 and 4 revealed that the interaction between sentences, pictures, and experiments was not significant (*estimate* = −0.003, *SE* = 0.002, $t$ = −1.557, $p$ = .121). Importantly, in both analyses, there was a main effect of Experiment ($t > 2$), which suggests that there were differences between experimental settings (e.g., "yes" vs. "no" correct response) of these studies, and, consequently, the results should be interpreted with caution. Collectively, these data support the conclusion that comprehenders integrate information from an object and its background, but the data are less strong for concluding that a null result provides evidence for visual simulation rather than amodally represented meaning. Thus, additional exploratory analyses were performed.

## Exploratory follow-up analyses

We argued that a low degree of semantic similarity between a scene from the sentence and that from the picture is one of the reasons why simulating, for example, a *sunlit bench* during reading should not work as a distractor when then verifying the picture of a *sunlit horse*. However, if this is the case, then the reverse should be true for scenes with higher degrees of semantic relatedness (e.g., *sunlit rose vs. sunlit scissors*). By contrast, in line with the amodal hypothesis, background information should be represented independently from the objects and thus always serve as a distractor leading to longer RTs. To address this issue, we computed the semantic similarity between sentence and picture scenes by using the University of Colorado's LSA@CU Boulder system (see Fig. 4, for more details) and then ran the same model as before, except that it included the "semantic similarity" predictor. There was a trending three-way interaction between sentences, pictures, and semantic similarity (*estimate* = 0.053, *SE* = 0.028, $t$ = 1.860, $p$ = .063). As shown in Fig. 4, longer RTs were observed for pictures with matching lighting information only when the semantic similarity between sentence and picture objects was high. A simple slopes analysis showed that this mostly occurred because participants were quicker to verify pictures with mismatching lighting information. This is particularly evident when looking at the results for sunlit pictures (*estimate* = −0.115, *SE* = 0.060, $t$ = −1.920, $p$ = .055) rather than moonlit pictures (*estimate* = −0.067, *SE* = 0.064, $t$ = −1.051, $p$ = .293).

Thus, the simulation hypothesis provides better support for the data from Experiment 4 as longer reaction times were observed only for more semantically related objects and not all matching sentence context-picture pairs. Furthermore, the data suggest that the locus of observed simulation effect is likely on sunlit and moonlit target objects rather than the lighting representation of the background. If this were not the case, then the lighting representation of the background would likely work as a distractor, regardless of the object being displayed Fig. 5.

[11] For this analysis, we again used the formula: *log.RT ~ sentence * picture * experiment + (1 + sentence * picture | ppt) + (1 | item).*

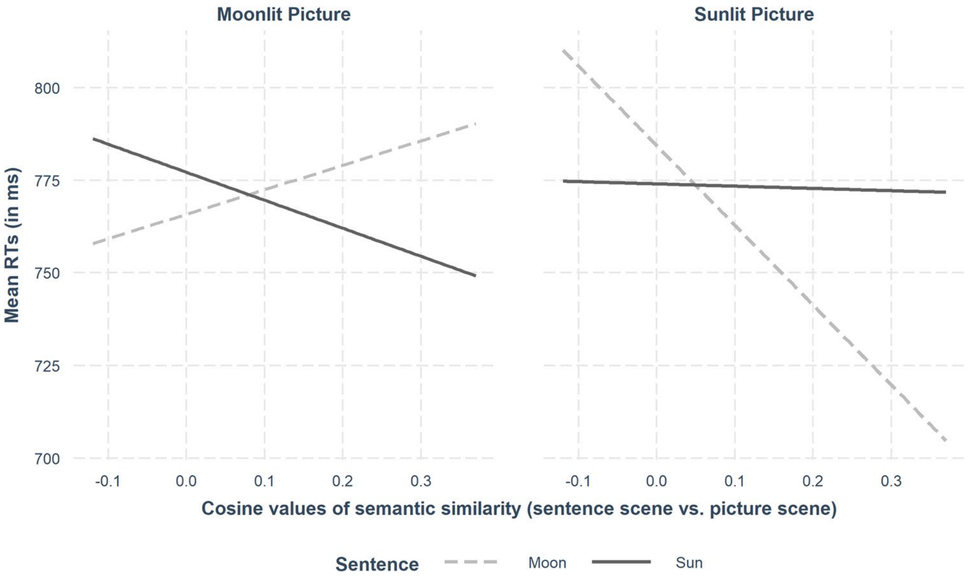

**Fig. 5** Mean response times (RTs) as a function of semantic similarity between objects (Experiment 4). *Note*. Mean cosine value was very low ($M$ = 0.04; $SD$ = 0.10), suggesting that most sentence-picture pairs from Experiment 4 had a low semantic similarity. Higher cosine values on the x-axis indicate a higher semantic similarity. The similarity between sentence and picture scenes was determined using the University of Colorado's Latent Semantic Analysis@CU Boulder system (document to document comparison type) that computes a cosine similarity score between -1 and 1 for each pair of terms (http://lsa.colorado.edu). The semantic similarity score for each sentence-picture condition was computed using adjectives describing a light setting and nouns referring to target objects (e.g., sunlit scissors vs. sunlit rose; sunlit scissors vs. moonlit rose; moonlit scissors vs. sunlit rose; and moonlit scissors vs. moonlit rose). According to this system, each word's representation is tantamount to a vector in the semantic space that summarizes the data about contexts in which that word is mentioned. Hence, the similarity between two texts is computed from the cosine between their vectors (see Landauer & Dumais, 1997, for more information on Latent Semantic Analysis)

## General discussion

In Experiments 1 and 3, background settings implied by the sentence influenced the speed with which participants verified pictured objects, such that responding was faster when both object and setting information from the picture matched sentence content. In contrast, in Experiment 2, the same background settings failed to influence the speed of responding when the processing of the sentence was interfered by visual noise. In Experiment 4, the same background settings had no effect on the speed of responses when the object presented for verification mismatched that mentioned in the sentence. This pattern of results suggests that language processing about objects and background settings relies on visual simulation. These findings support theories of grounded cognition that posit that language comprehension invokes perceptual symbols in the simulation of described events (Barsalou, 1999, 2008; Glenberg & Robertson, 1999; Zwaan, 2004). Furthermore, these findings are consistent with other empirical evidence on the importance of background settings for conceptual processing (e.g., Horton & Rapp, 2003; Wu & Barsalou, 2009; Yaxley & Zwaan, 2007).

Our results are hard to accommodate by the account of backward semantic priming, which suggests that knowledge is represented in an amodal format. While this account also predicts a congruency effect for both versions of the light source, it does not predict the elimination of the difference between matching and mismatching conditions when the visual simulation is disrupted by visual noise (as demonstrated in Experiment 2). Similarly, it does not predict that verification times of sunlit and moonlit scenes with non-present objects should be unaffected after reading the semantically related "sun" and "moon" words, respectively. That RTs remained the same for trials with matching background settings but mismatching target objects is consistent with the view that entities and situations become active together in the simulation process (Barsalou, 2005).

It is perhaps remarkable that there was no suggestion that visual interference affected participants' verification times for matching sentence-picture pairs in Experiment 2 compared to Experiment 1. Indeed, visual interference in Experiment 2 only reduced RTs of mismatching sentence-picture pairs, thereby suggesting that visual noise disrupted incongruent visual content activated by listening to a sentence, content that otherwise hinders verifying the picture. On the one hand, these results may seem surprising in light of Ostarek et al.'s (2019) results regarding shape simulation, where longer RTs for trials with visual interference were reported. On the other hand, these results are less surprising when placed alongside evidence reported by Edmiston and

Lupyan (2017) on judging the orientation of objects, where visual noise led to faster RTs on mismatching-object trials (e.g., verifying an upright picture of an alligator after hearing "dog"). Thus, it looks like the effect of visual interference on visual simulation is rather specific and depends on the type of content being simulated.

The findings of the present research suggest that determining in exactly what situations visual simulations are more important than amodal representations may lead to more valuable insights than determining whether the results are merely consistent with an embodied account or not (see also Ostarek & Bottini, 2021, for a related discussion). Specifically, our results from Experiment 4 point to the tentative conclusion that the language system may suffice to understand events when semantic consistency between objects and their backgrounds is high (e.g., sunlit rose vs. sunlit scissors). For a deeper understanding of semantically inconsistent events, which made the bulk of the present research (sunlit bench vs. sunlit horse), relying on the simulation system is necessary.

In conclusion, the present research makes two contributions to the literature. First, it shows that comprehenders create the experience of "being there in the scene" via integrated simulation of both target objects and background settings. Second, previous studies demonstrating the causal role of visual processes for language processing and object knowledge have primarily focused on object properties (e.g., Davis et al., 2020; Edmiston & Lupyan, 2017; Ostarek & Huettig, 2017; Rey et al., 2017); the present study demonstrates that background information is also represented in a visual format.

## Appendix 1

Accuracy scores and response times for Experiments 1–4

| | Dependent variable | | | |
|---|---|---|---|---|
| | Accuracy | | RT | |
| | Moonlit Picture *M (SD)* | Sunlit Picture *M (SD)* | Moonlit Picture *M (SD)* | Sunlit Picture *M (SD)* |
| Experiment 1 | | | | |
| Moon Sentence | 0.98 (0.14) | 0.99 (0.12) | 667 (260) | 694 (273) |
| Sun Sentence | 0.98 (0.14) | 0.98 (0.13) | 696 (280) | 651 (241) |
| Experiment 2 | | | | |
| Moon Sentence | 0.98 (0.15) | 0.97 (0.16) | 668 (251) | 669 (252) |
| Sun Sentence | 0.98 (0.14) | 0.97 (0.16) | 663 (245) | 671 (261) |
| Experiment 3 | | | | |
| Moon Sentence | 0.98 (0.15) | 0.97 (0.16) | 681 (253) | 723 (292) |
| Sun Sentence | 0.97 (0.17) | 0.98 (0.15) | 710 (275) | 694 (267) |
| Experiment 4 | | | | |
| Moon Sentence | 0.98 (0.13) | 0.98 (0.13) | 759 (285) | 769 (285) |
| Sun Sentence | 0.98 (0.15) | 0.98 (0.14) | 764 (280) | 762 (293) |

*Note. M* = mean, *SD* = standard deviation. Participants with an accuracy threshold of 80% or higher were included in the analysis. Mean response times (RT) were calculated using correct responses only

**Acknowledgements** This research was supported by the Portuguese Foundation for Science and Technology with a grant SFRH/BPD/115533/2016 awarded to the first author.

**Code availability** The code is available at the website of the Open Science Framework https://osf.io/xsrz4/?view_only=dc6836edd3e548bcb032de2293ae0bcf

**Authors' 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.

**Funding** This research was supported by the Portuguese Foundation of Science and Technology with a grant SFRH/BPD/115533/2016 awarded to the first author.

**Data availability** The data and materials are available at the website of the Open Science Framework https://osf.io/xsrz4/?view_only=dc6836edd3e548bcb032de2293ae0bcf

## Declarations

**Competing interests** The authors declare that no competing interests exist.

**Ethics approval** All experiments were carried out in accordance with the World Medical Association's Declaration of Helsinki and the ethical guidelines of the host institution.

**Consent to participate** Informed consent was obtained from all participants prior to their participation.

**Consent for publication** We grant the Publisher the right to publish this research in the event of acceptance.

## References

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

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. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Barsalou, L. (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. (2005). Situated conceptualization. In: H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 619–650). Elsevier.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2016). Situated conceptualization: Theory and applications. In: *Foundations of embodied cognition: Perceptual and emotional embodiment* (pp. 11–37). Routledge/Taylor & Francis Group.

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. https://doi.org/10.18637/jss.v067.i01.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*, 77–80. https://doi.org/10.1126/science.177.4043.77

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Concepcion, R. (2019). *Adobe Photoshop and Lightroom Classic CC Classroom in a Book (2019 release)*. Adobe Press.

Connell, L. (2007). Representing object colour in language comprehension. *Cognition*, *102*(3), 476–485. https://doi.org/10.1016/j.cognition.2006.02.009

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x

Davis, C. P., Joergensen, G. H., Boddy, P., Dowling, C., & Yee, E. (2020). Making it harder to "see" meaning: The more you see something, the more its conceptual representation is susceptible to visual interference. *Psychological Science*, *31*(5), 505-517. https://doi.org/10.1177/0956797620910748

de Koning, B. B., Wassenburg, S. I., Bos, L. T. & van der Schoot, M. (2017. Mental simulation of four visual object properties: similarities and differences as assessed by the sentence-picture verification task. *Journal of Cognitive Psychology*, 29, 420-432. https://doi.org/10.1080/20445911.2017.1281283

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

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, *28*(1), 1–26. https://doi.org/10.1080/01638539909545067

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. https://doi.org/10.1111/cogs.12870

Horchak, O. V., & Garrido, M. V. (2021). Dropping bowling balls on tomatoes: Representations of object state-changes during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(5), 838–857. https://doi.org/10.1037/xlm0000980

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

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

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

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2018). Mixedpower: a library for estimating simulation-based power for mixed models in R. https://doi.org/10.5281/zenodo.1341047

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01546-0

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. https://doi.org/10.18637/jss.v082.i13

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211. https://doi.org/10.1037/0033-295X.104.2.211

Latent Semantic Analysis @ CU Boulder. (n.d.) University of Colorado. Available at: http://lsa.colorado.edu/. Accessed 5 October 2021.

Masson, M. E. J. (2015). Toward a deeper understanding of embodiment. *Canadian Journal of Experimental Psychology = Revue Canadienne De Psychologie Experimentale*, *69*(2), 159–164. https://doi.org/10.1037/cep0000055

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *15*(6), 1003–1019. https://doi.org/10.1037//0278-7393.15.6.1003

Neill, W. T., & Valdes, L. A. (1992). Persistence of negative priming: Steady state or decay? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 565–576. https://doi.org/10.1037/0278-7393.18.3.565

Ostarek, M., & Bottini, R. (2021). Towards strong inference in research on embodiment – Possibilities and limitations of causal paradigms. *Journal of Cognition*, *4*(1), 5. https://doi.org/10.5334/joc.139

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

Potter, M. C. (1975). Meaning in visual search. *Science*, *187*(4180), 965–966. https://doi.org/10.1126/science.1145183

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Rey, A. E., Riou, B., Vallet, G. T., & Versace, R. (2017). The automatic visual simulation of words: A memory reactivated mask slows down conceptual access. *Canadian Journal of Experimental Psychology, 71*(1), 14–22. https://doi.org/10.1037/cep0000100

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, *51*(3), 1258–1270. https://doi.org/10.3758/s13428-018-1099-3

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

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, *61*(6), 1140–1153. https://doi.org/10.3758/BF03207619

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

Winter, B., & Bergen, B. (2012). Language comprehenders represent object distance both visually and auditorily. *Language and Cognition*, *4*(1), 1–16. https://doi.org/10.1515/langcog-2012-0001

Wu, L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*(2), 173–189. https://doi.org/10.1016/j.actpsy.2009.02.002

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

Yuval-Greenberg, S., & Heeger, D. J. (2013). Continuous flash suppression modulates cortical activity in early visual cortex. *Journal of Neuroscience*, 33(23), 9635–9643. https://doi.org/10.1523/JNEUROSCI.4612-12.2013

Zwaan, R. A. (2004). The Immersed Experiencer: Toward an Embodied Theory of Language Comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory*, (vol. 44, pp. 35–62). Elsevier Science.

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., 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

**Open Practices Statement** The data and materials for all experiments reported here are available at the Open Science Framework website: https://osf.io/xsrz4/?view_only=dc6836edd3e548bcb032de2293ae0bcf. None of the experiments was preregistered.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.