Changes in scene distance automatically drive scaling of object representations

Giacomo Aldegheri[1,2,3]*, Surya Gayet[1,4] & Marius V. Peelen[1]

[1] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

[2] Department of Experimental Psychology, Justus-Liebig University, Gießen, Germany

[3] Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig University, Gießen, Germany

[4] Helmholtz Institute, Experimental Psychology, Utrecht University, Utrecht, The Netherlands

* Corresponding author: giacomo.aldegheri@gmail.com

Abstract

As we navigate our visual environment, our viewpoint shifts, causing predictable changes in object appearance. Moving forward, for instance, increases the retinal size of objects in a scene, proportionally with the distance travelled. Such regularities can be exploited to predict visual object transformations, thereby facilitating object perception. Previous research showed that observers automatically predict the orientation of an object from the rotation of the surrounding scene. It remains unknown, however, whether this is a ubiquitous property of human vision that generalizes to other transformations. In three behavioral experiments (N=151), we investigated whether observers automatically predict the retinal size of temporarily occluded objects during forward motion. Participants performed a perceptual discrimination task on the objects that reappeared in a size that either matched or mismatched the change in viewing distance conveyed by the scene context. We found that scene-driven size expectations strongly influenced task performance. This effect remained consistent even when size expectations were violated on a majority of trials, suggesting that scene context elicits automatic predictions that cannot easily be overruled by short-term evidence. We conclude that scenes drive predictions of object transformations, capitalizing on the predictable ways in which visual input is altered when our viewpoint changes.

Keywords: Scene perception, Mental transformations, Visual expectations, Object perception

Introduction

As we move through our environment, the visual input evoked by objects in the world systematically changes along with our changing viewpoint. For instance, when we walk around an object, we see it rotating through the depth plane, and as we get closer or farther, the retinal size of an object increases or decreases. Predicting how these transformations affect the appearance of objects may allow us to track, detect, or recognize objects more readily, and is therefore crucial for navigating the world effectively. The task of predicting object transformations is computationally challenging, however, as it requires several inferences to be made. In the case of mental rotation, for example, the direction and amount of rotation to apply to the object must be determined (Hamrick & Griffiths, 2014). It would be advantageous, then, if we could exploit the redundancy of real-world scenes by mentally transforming objects coherently with their context.

For at least some mental transformations, this seems to be the case: in our recent work, we found that object representations are automatically rotated concurrently with the viewpoint of the surrounding scene (Aldegheri et al., 2023, 2025). It remains unknown, however, whether this influence of scene context on mental object transformations is a general phenomenon or whether it is limited to rotation. Mental rotation has been studied extensively (Shepard & Metzler, 1971; Cooper & Shepard, 1973; Shepard & Cooper, 1982; Larsen, 2014; Xue et al., 2017), but prior research indicates that it is merely one among many transformations that we are able to simulate in our minds. For example, humans are able to mentally *translate* or *scale* objects similarly to how they rotate them (Bennett, 2002; Bundesen et al., 1983; Bundesen & Larsen, 1975; Larsen & Bundesen, 1978, 1998; Schmidt et al., 2016; Sekuler & Nash, 1972). This suggests that predicting object transformations might be a generalized cognitive capacity, encompassing several different ways in which the visual input from objects can change as we navigate the world. Because the way in which object properties change may be predicted from changes in scene context (as described above), we asked whether changes in scene context are used to predict object transformations, beyond rotation alone. This would indicate that predicting mental object transformations from changes in scene context is a principled way in which we exploit redundancies in the environment to reduce computational cost.

Here, to investigate the role of scene context in driving mental transformations beyond rotation, we focus on *scaling*, the predictable change in an object's retinal size as a function of viewing distance. As objects rarely physically shrink or expand in the real world, their retinal size mostly varies with our distance from them: accordingly, behavioral evidence suggests that we generally perceive size changes as translations in depth (Bundesen et al., 1983; Larsen & Bundesen, 2009). Because retinal size depends on distance, scene context should play a crucial role in influencing our representations of object size, as real-world scenes contain a rich variety of depth cues (Landy et al., 1995). Indeed, the perceived size of an object has long been known to be altered by pictorial depth cues in a scene, such that objects farther away are perceived as larger, reflecting their inferred real-world size (Leibowitz et al., 1969; Yildiz et al., 2021; Yeh et al., 2024). The influence of scene context on object size is not limited to perceived objects: it can also affect internally generated representations of objects that are not (yet) perceived. When observers search for objects at particular distances in a scene, their *search templates* - the top-down object representations evoked during visual search - have been shown to scale in accordance with search distance (Gayet & Peelen, 2022; Gayet et al., 2024). Internal representations of objects, then, can be adaptively rescaled based on the scene context. A still open question, however, is whether the rescaling of object representations occurs dynamically, as the distance to the scene changes. If visual object representations would dynamically update in accordance with changes in the surrounding scene context (e.g., through rotation and scaling), this would facilitate the perception of, and interaction with, objects in dynamic real-world environments.

We ran a series of online behavioral experiments to determine whether internal object representations are rescaled according to scene viewpoint. To this end, we built on an experimental paradigm that we recently developed to study changes in object rotation (Aldegheri et al., 2023). On each trial, we showed participants an object placed within a realistic 3D scene (**Figure 1**). The viewpoint on the scene translated in depth, with the camera 'zooming in'. During

this shift in viewpoint, the object was concealed by an occluder. Eventually, the object reappeared, either with a size consistent with the new viewing distance on the surrounding scene (Congruent trials), or with an inconsistent size (Incongruent trials). Importantly, the amount of scene translation (small or large) varied randomly across trials, always in the same number of discrete steps while the object was occluded (**Figure 2**), so that whether the object had been rescaled congruently or incongruently could only be known by integrating the change in scene distance and the change in object size, both relative to the original scene, at the start of the trial (before occlusion). In other words, neither the final snapshot alone nor the amount of scene translation alone provided information about object size congruency. Participants had to perform an orthogonal visual discrimination task on the object that reappeared, judging whether two versions of the object were the same or different. Importantly, neither the scene context, nor the expectation of object size that the scene may convey, were relevant for the discrimination task. We compared performance on this discrimination task between Congruent and Incongruent trials, and found that expectations of object size, driven by the scene, substantially influenced participants' responses. This suggests that scene context drives internal predictions of object size even when this is not required for the task at hand. Moreover, across three experiments, we manipulated the probability that the objects reappeared in a Congruent size (i.e., in accordance with the change in scene viewpoint). We found that even when scene-driven expectations were violated on a large proportion of trials, they still influenced behavioral responses in a similar manner, showcasing the obligatory and automatic nature of the influence of scene context on object transformations. Altogether, these results indicate that mental scaling can be driven by scene context in an automatic way, analogously to rotation, pointing to a general role of scene viewpoint in driving transformations of object representations.

Methods

Participants

All experiments were run online, hosted on Pavlovia (https://pavlovia.org/) and programmed in Javascript using JsPsych 6.3.0 (De Leeuw, 2015) and the jspsych-psychophysics library (Kuroki, 2021). Participants were recruited on Prolific (Palan & Schitter, 2018) and had to meet the following criteria: reside in Europe or the UK, to ensure that they were participating during day hours; have participated in at least 10 previous studies on Prolific; and have a Prolific approval rate of at least 95%.

Participants provided informed consent before the study and received monetary compensation for their participation. The study was approved by the Radboud University Faculty of Social Sciences Ethics Committee (ECSW2017-2306-517). Participants were included in the analysis if a one-sided binomial test comparing their accuracy in our same/different task with 50% was significant (at $\alpha = 0.05$), meaning that they were performing better than chance across all conditions. We continued data collection until the number of included participants reached 50 for each experiment. In Experiment 1, we excluded 47 participants. Of the included 50 participants, 24 were female, 25 were male, and one participant's demographic information was missing. Mean age was 27.1 ± 4.1. In Experiment 2, we excluded 37 participants. Of the included 50 participants,

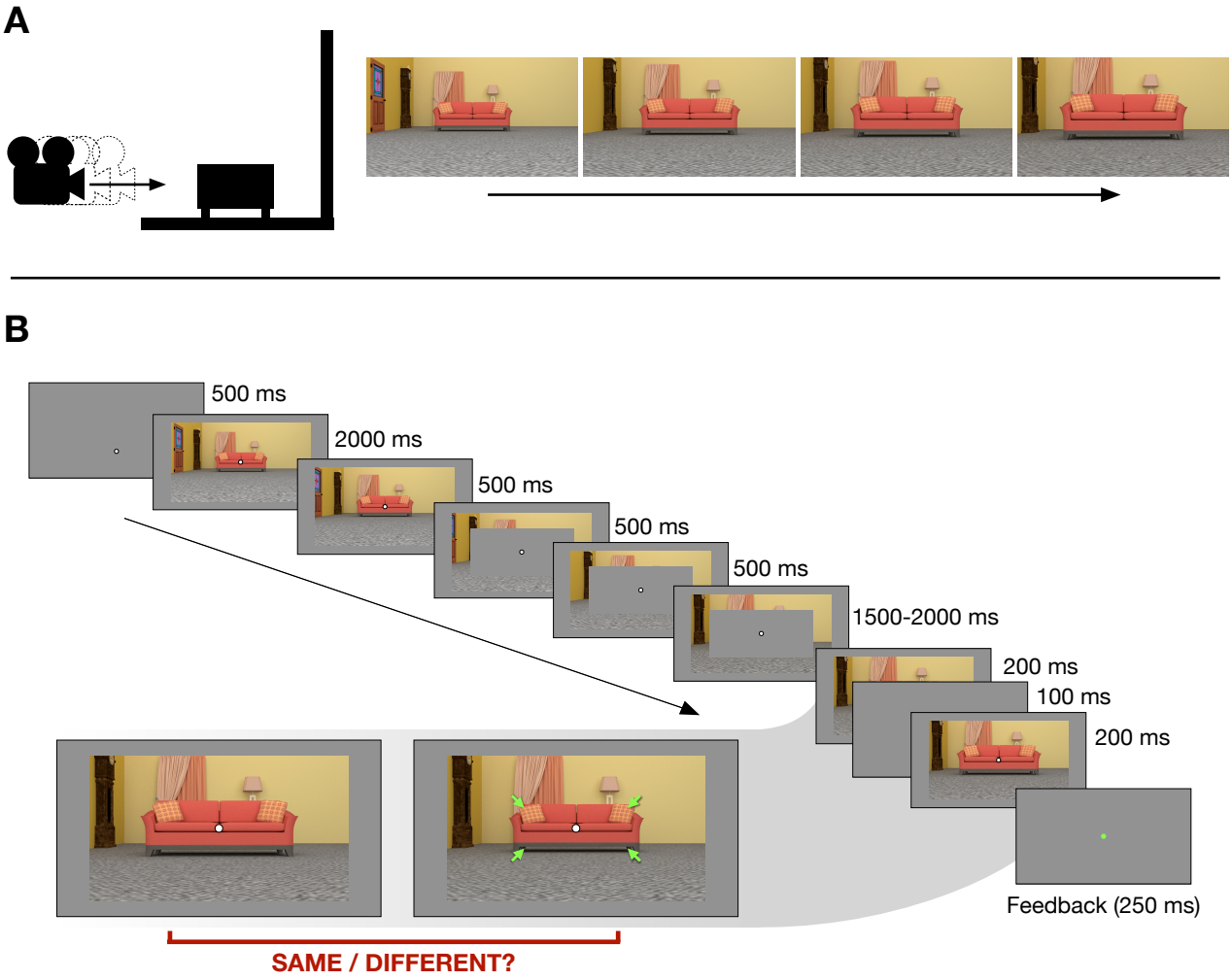

Figure 1: (A) Illustration of the viewpoint shift sequences that were shown during the experiment. A realistic 3D scene featuring a central object was shown, and viewpoint sequences were generated by gradually moving the camera closer to the object with a constant speed. Discrete snapshots of these sequences were shown in the experiment. **(B)** Example of a trial: after the object was shown in its initial position, the scene viewpoint started shifting toward the object ('zooming in'). During most of this viewpoint shift, the object was occluded by a grey rectangle, and after the shift was completed, the occluder disappeared, revealing the object. The object was briefly flashed twice (200 ms each, with a 100 ms ISI), either with the exact same position (and retinal size), or two slightly different positions. Participants' task was to judge whether the two appearances of the object looked the 'same' or 'different'. In this example, they are different: the object is slightly smaller in the second appearance (arrows added for illustration).

25 were female, 24 were male, and 1 participant's information was missing. Mean age was 25.8 ± 4.8. In Experiment 3, we excluded 42 participants. Of the included 51 participants, 21 were female, and mean age was 26.7 ± 4.5.

The high exclusion rate was likely due to the difficulty of the task. The difference between probe stimuli was defined in 3D space (object position in depth), limiting the range of possible stimulus differences we could show. We wanted to make sure that the difference between different object positions (and thus Congruent and Incongruent positions) was noticeable, so probe objects could appear at either very near or very far distances. We could thus not use the full range of distances available in the scene, and for far object positions, depth differences were very hard to notice. Moreover, we kept the presentation time short (200 ms) for each of the two target stimuli, in order to reduce the influence of deliberate judgment and investigate how scene-

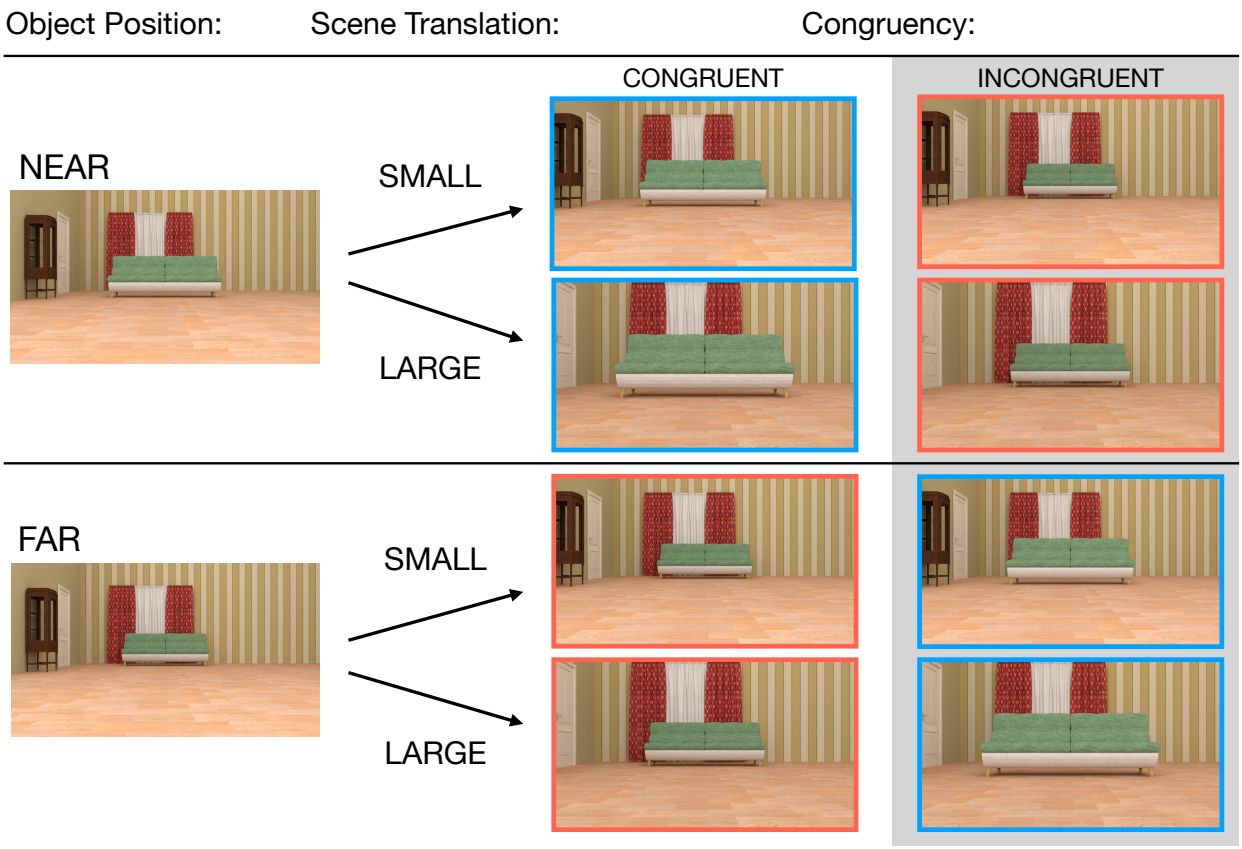

Figure 2: Illustration of the experimental design, showing the initial position of the object relative to the scene (either Near or Far), and the final images (after the whole scene translation sequence and the occlusion period) resulting from a Small or Large translation on Congruent or Incongruent trials. As highlighted by the color frames, the same (final) image could appear as either Congruent or Incongruent on different trials.

driven expectations influenced a primarily perceptual task. Importantly, however, all results reported here remained consistent with no participant exclusions (**Figure S1**).

Stimuli

The stimuli were based on 4 different indoor scenes (**Figure S2**) modeled in Blender 2.92 (Blender Foundation) and rendered using the Cycles rendering engine for realistic lighting. The scenes all had the same layout (floor, two walls at a right angle and a main object in the center) but the main object varied, as well as the objects present in the background (adjacent to the walls), and the textures on the walls and floor. The central object could be a couch or a bed: we chose large, immovable object categories that are generally expected to remain in a fixed position within a scene. For each scene, a sequence of different viewpoints was rendered by translating the camera gradually closer to the scene (zooming in, **Figure 1A**). The main object was always fully included in the frame, while other background objects could go out of the frame. The main object was always presented with its longer side (front for the couch, side for the bed) facing the viewer. The height and pitch of the camera were chosen so that the main object would always remain at the center of the scene. All scene images had a resolution of 960 x 540 pixels.

Procedure

Each trial (**Figure 1B**) began with a fixation dot (which was always present during the trial, radius 5 pixels) for 500 ms, followed by the first view of the scene for 2000 ms, the 3 intermediate views for 500 ms each, and the final view for a randomly jittered duration between 1500 and 2000 ms. The central object (couch or bed) was fully visible for the first and second view, and was occluded

by a grey rectangle during the third, fourth and final view. The occluder had the height and width of the largest possible view of the object in a specific scene, plus a margin (horizontal margin: 110 pixels, vertical: 40 pixels) to ensure the object was fully covered and its shadow was not visible, which would have provided a cue to its size behind the occluder.

After the final view of the scene was shown, the occluder disappeared, briefly revealing the object (within the scene) twice, for 200 ms each, with a 100 ms inter-stimulus interval. We will refer to these two brief presentations of the object as the *probes*. Participants' task was to report whether the second probe was the 'same' as, or 'different' from, the first, by pressing the F or J key, respectively. After responding, they would receive feedback: the fixation dot would turn green following a correct answer, and red following an incorrect one, for 250 ms. They had a maximum of 2500 ms to respond, after which the fixation dot would turn black, the experiment would skip to the next trial and the current trial would be counted as missed.

Participants were explicitly told that their task would be on the final two snapshots of the objects exclusively. They were additionally instructed to nonetheless pay attention to the preceding sequence of images. The position of the two probes in depth, and thus their retinal size, was defined in terms of distance in the virtual scene, using the default Blender unit. We henceforth refer to this measurement unit as *arbitrary unit* (AU). The first probe was randomly sampled from a normal distribution (SD = 0.05 AU) centered around the Congruent or Incongruent object position, to add a small amount of jitter, and then rounded to show the nearest rendered view (views were rendered in steps of 0.025 AU). On half of trials, the second probe was exactly the same as the first probe ('same' trials), thus requiring a 'same' response. On the other half of trials ('different' trials), the probe object was translated in depth relative to the first (see **Figure 1**, bottom left), thus requiring a 'different' response. The shift could occur in forward or backward direction with equal probability.

The depth difference between the two probes (on 'different' trials) was titrated using a 2-down 1-up staircase, to keep the task difficulty constant across participants and across experiments. Specifically, a single staircase was used across both Congruent and Incongruent trials to ensure overall performance was around 70% correct (Wetherill & Levitt, 1965) across conditions, while still allowing for accuracy differences between the Congruent and Incongruent conditions. The depth difference was adjusted after both 'same' and 'different' trials. The starting value for the staircase was 1 AU, the initial step size was 0.05 AU (but was halved after 3 staircase reversals), and the minimum and maximum possible depth differences shown were 0.025 and 1 AU, respectively. The means and standard deviations of the depth differences reached by the staircase in the second half of trials were 0.68 ± 0.23 in Experiment 1, 0.72 ± 0.19 in Experiment 2, and 0.7 ± 0.20 AU in Experiment 3.

Each experiment lasted about 30 minutes in total and comprised 8 experimental blocks, after each of which participants were encouraged to take a short break. Before the experiment began, participants read the on-screen instructions, accompanied by demonstration images, at their own pace. Then they completed a short practice run. During the practice run, the presentation time of the two probes gradually decreased across trials, from 300 ms to the eventual presentation time that was used in the main experiment (200 ms). This allowed participants to familiarize with the task with an initially less challenging presentation time.

Experimental design

Trials varied along four factors (**Figure 2**): Congruency (Congruent, Incongruent), initial Object Position relative to the scene (two possible distances from the observer, Near or Far), amount of Scene Translation (Small or Large), and Scene (1 of 4 different exemplars). The only difference between the three experiments is that the proportion of Congruent and Incongruent trials was varied (75% of total trials were Congruent in Exp. 1, 50% in Exp. 2, and 25% in Exp. 3).

All factors of non-interest (2 x Object Position, 2 x Scene Translation, and 4 x Scene) were fully balanced within Congruent and Incongruent trials. The experiment was divided into four partitions (assigned to be either Congruent or Incongruent depending on the experiment), in which each of these (2 x 2 x 4 =) 16 conditions of non-interest were repeated 3 times. This resulted in a total of 192 trials in each experiment. All these trials were presented in randomized order throughout the experiment.

On Incongruent trials, the object reappeared at the end of the sequence in a position that was inconsistent with the position of the object shown at the start of the sequence: on Near trials, the object appeared in the Far position, and vice versa. By doing so, we ensured that the exact same images were used as Congruent in the context of one trial, and Incongruent in another, avoiding any possible confounds due to visual differences between conditions (**Figure 2**).

Data analysis

We used three different measures of performance: (1) accuracy (percentage of correct responses over all trials), (2) sensitivity (d') and (3) criterion. All three measures were computed separately for each condition of interest (Congruent and Incongruent trials). We computed d' and criterion in order to disentangle the influence of scene-driven expectations on observers' sensitivity and bias. We consider 'Same' trials as noise, and 'Different' as signal, meaning that a positive criterion indicates a bias towards responding 'same'. We used the loglinear method (Hautus, 1995) to correct for the rare cases of 100% accuracy in a particular condition. Sensitivity (d') and criterion (c) were thus computed as follows:

$$d' = \Phi^{-1}(H_{corr}) - \Phi^{-1}(FA_{corr})$$

$$c = -\frac{1}{2}(\Phi^{-1}(H_{corr}) + \Phi^{-1}(FA_{corr}))$$

Where Φ^{-1} is the inverse of the normal cumulative distribution function, and H_{corr} and FA_{corr} are the loglinear-corrected hit and false alarm rates, respectively, which were obtained as follows:

$$H_{corr} = \frac{N_{hit} + 0.5}{N_{hit} + N_{miss} + 1}$$

$$FA_{corr} = \frac{N_{FA} + 0.5}{N_{FA} + N_{CR} + 1}$$

Where N_{hit}, N_{miss} are the numbers of hit and miss trials, and N_{FA} and N_{CR} are the numbers of false alarm and correct rejection trials, respectively. Trials on which participants did not provide a response were excluded from the analysis.

Post-experiment survey

After completing the experiment, participants were asked three questions that would help us gauge their awareness of the key experimental manipulation. The questions were:

- "Your task was only on the final image, when the object changed or not. Did you also pay attention to the sequence of images before the task image?" - the response had to be indicated on a Likert scale from 1 (Not at all) to 7 (All the time).
- "When the scene shifted, did you anticipate seeing the object in the correct viewpoint after it reappeared?" - the response also had to be indicated on a 1-7 Likert scale.
- "What percentage of objects were in line with your expectation? (They reappeared with the correct viewpoint)" - the response had to be a value in percentage, from 0 to 100%.

Analysis software

All analyses were conducted in Python using Pandas 1.2.5 (McKinney, 2011), Numpy 1.20.2 (Harris et al., 2020), Pingouin 0.3.4 (Vallat, 2018), and Scipy 1.6.2 (Virtanen et al., 2020), and results were visualized using Matplotlib 3.3.4 (Hunter, 2007), and Seaborn 0.11.1 (Waskom, 2021).

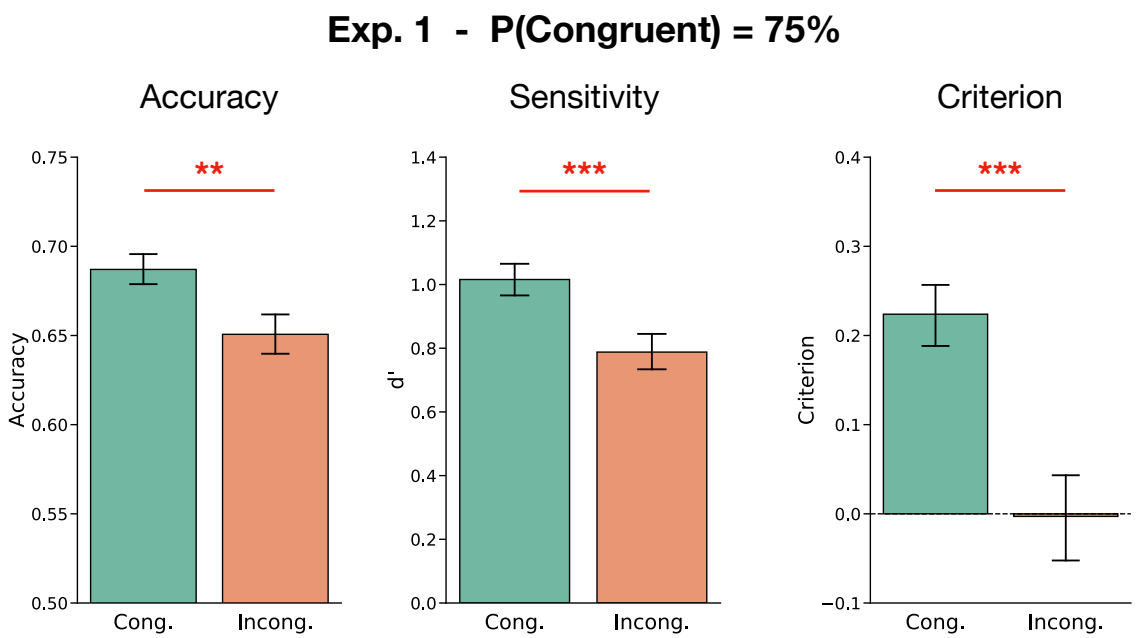

Figure 3: Results of Experiment 1 (75% Congruent trials). Mean (and SEM) accuracy, sensitivity, and criterion for the Congruent and Incongruent trials. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Results

Experiment 1: 75% of Congruent trials

In the first experiment (**Figure 3**), the object appeared in the congruent view (given the scene context) on a majority (75%) of trials. Across conditions, participants' mean accuracy (and SEM) was 0.68 ± 0.01, indicating that they were able to perform the task, and that the staircase successfully approached the desired accuracy of 70%. Analyzing overall criterion, we found that it was significantly above zero (mean: 0.17, $t(49) = 5.44$, $p < 0.001$, $d = 0.77$, 95% CI = [0.11, 0.23]), indicating a general bias towards responding 'same', possibly due to the relatively small perceptual differences between the probes.

In our key analyses, we first compared accuracy between the Congruent and Incongruent conditions. We found participants to be significantly more accurate on Congruent than Incongruent trials (means: 0.69 vs. 0.65; $t(49) = 3.17$, $p = 0.003$, $d = 0.53$, 95% CI = [0.01, 0.06]), as shown in **Figure 3**. Next, analyzing sensitivity and criterion separately, we found that both measures were significantly influenced by congruency (mean sensitivity: 1.02 vs. 0.79; $t(49) = 3.70$, $p < 0.001$, $d = 0.60$, 95% CI = [0.10, 0.35]; mean criterion: 0.22 vs. 0.00; $t(49) = 4.49$, $p < 0.001$, $d = 0.78$, 95% CI = [0.13, 0.33]). Participants were more sensitive when performing the task on objects with a congruent size. Additionally, they tended to respond 'different' more often for incongruent objects, canceling out their overall bias. Overall, this result indicates that scene-driven size expectations significantly affected participants' responses in the task, despite the absence of explicit requirements to predict object size.

Experiment 2: 50% of Congruent trials

In **Experiment 1**, the object reappeared with a size that matched participants' scene-driven expectations on a majority of trials. Thus, the short-term expectations established during the experiment matched the long-term expectations derived from real-world regularities (the fact that objects are transformed coherently with the surrounding scene). In **Experiment 2** (**Figure 4**), we investigated whether long-term (real-world) expectations would still affect task performance when congruent and incongruent object sizes appear with the same probability during the experiment (i.e., equating short-term expectations). This experiment was identical to **Experiment 1** in all

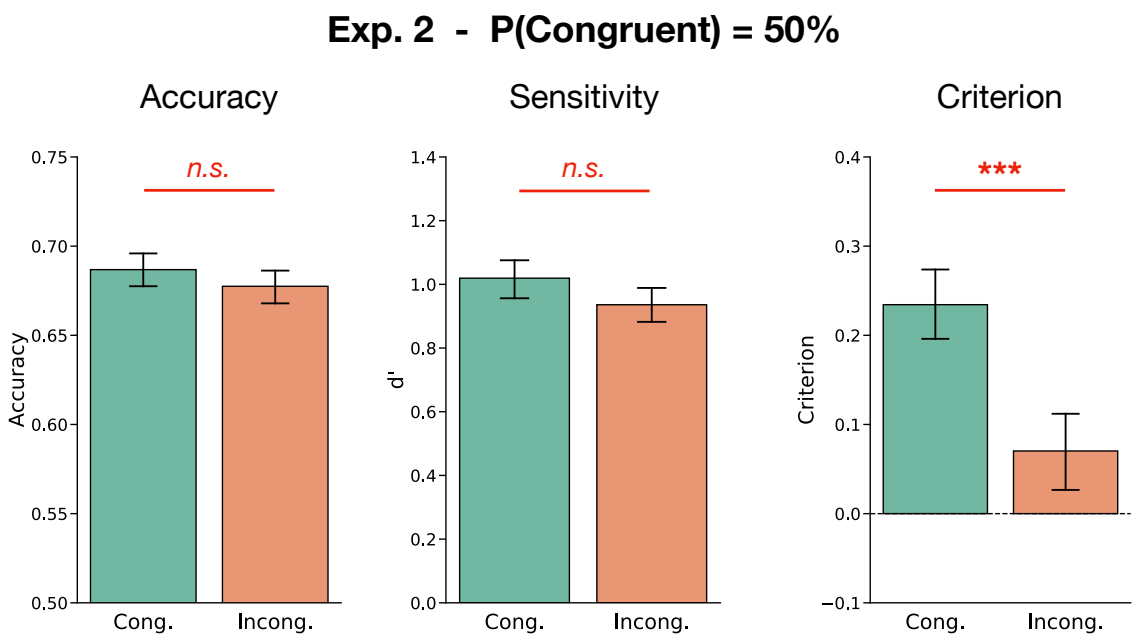

Figure 4: Results of Experiment 2 (50% Congruent trials). Mean (and SEM) accuracy, sensitivity, and criterion for the Congruent and Incongruent trials. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

respects, except that now objects reappeared from occlusion in a congruent size in 50% of trials (instead of 75%).

Like in **Experiment 1**, participants were able to perform the task well above chance level (mean accuracy and SEM: 0.68 ± 0.01). Also consistently with the previous experiment, their criterion was significantly above zero (mean: 0.15, $t(49) = 4.70$, $p < 0.001$, $d = 0.66$, 95% CI = [0.09, 0.22]), meaning that they had a bias towards reporting 'same' (i.e., no change between the two probes).

Comparing accuracy between the Congruent and Incongruent conditions, we found no significant difference (means: 0.69 vs. 0.68; $t(49) = 0.89$, $p = 0.375$, $d = 0.14$, 95% CI = [-0.01, 0.03]). Analyzing the effects on sensitivity and criterion separately, we found no significant difference of congruency on sensitivity (means: 1.02 vs. 0.94; $t(49) = 1.34$, $p = 0.187$, $d = 0.21$, 95% CI = [-0.04, 0.21]). On the other hand, criterion was significantly higher for congruent than incongruent objects (means: 0.23 vs. 0.07; $t(49) = 3.07$, $p = 0.003$, $d = 0.55$, 95% CI = [0.06, 0.27]). This difference between conditions entails that participants were still forming an expectation of the object size, as implied by the scene, and that this expectation was influencing their responses in the task. Unlike in **Experiment 1**, however, there was no difference in perceptual sensitivity when objects reappeared in congruent or incongruent sizes with equal probability. We address possible reasons for this discrepancy in the **Discussion**.

Experiment 3: 25% of Congruent trials

In **Experiment 3** (**Figure 5**), we asked whether scene-driven predictions of object size would reverse if the object reappeared in a congruent size on a minority (i.e., 25%) of trials. In this situation, the scene context is *counter*predictive of the object size. If effects of scene context remain consistent with those of Experiments 2 and 3 (rather than reversing), this would testify to the automaticity of scene-driven predictions of object size.

Consistently with the previous experiments, participants performed the task well above chance level (mean accuracy and SEM: 0.69 ± 0.01). Also consistently with the previous experiments, they showed a strong overall bias towards responding 'same', leading to a significantly positive criterion (mean: 0.13, $t(50) = 4.10$, $p < 0.001$, $d = 0.57$, 95% CI = [0.06, 0.19]).

Comparing Congruent and Incongruent trials, we found no significant difference in accuracy (means: 0.70 vs. 0.68; $t(50) = 1.16$, $p = 0.253$, $d = 0.17$, 95% CI = [-0.01, 0.03]) and no significant difference in sensitivity (means: 1.07 vs. 0.99, $t(50) = 1.20$, $p = 0.236$, $d = 0.18$, 95% CI

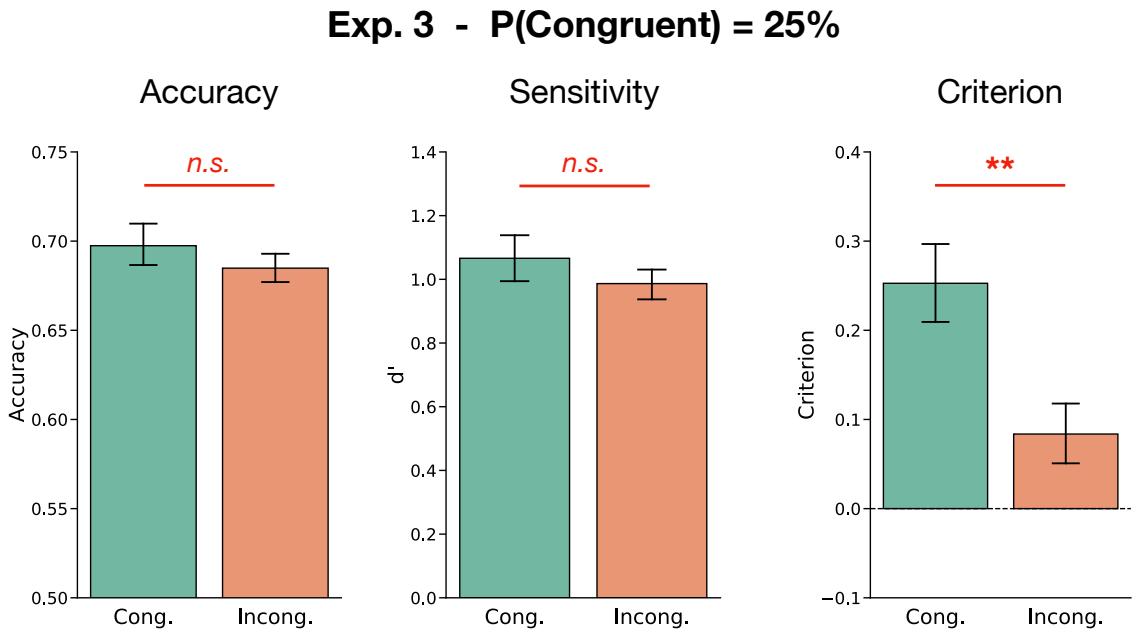

Figure 5: Results of Experiment 3 (25% Congruent trials). Mean (and SEM) accuracy, sensitivity, and criterion for the Congruent and Incongruent trials. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

= [-0.05, 0.21]). On the other hand, criterion was significantly higher for Congruent than Incongruent trials (means: 0.25 vs. 0.08, $t(50) = 3.68$, $p < 0.001$, $d = 0.61$, 95% CI = [0.08, 0.26]). Thus, object size congruency influenced participants' responses in a manner that was qualitatively similar to Experiment 1, even though scene-based expectations were now violated on a majority of trials.

Unlike **Experiment 1** (but similarly to **Experiment 2**), however, this shift in criterion was not accompanied by an effect on perceptual sensitivity. Importantly, the numerical direction of all results (differences in accuracy, sensitivity, and criterion) was the same as in the previous two experiments, indicating that short-term violations of real-world regularities did not reverse their effect on performance.

Congruency-probability interaction

Analyzing the three experiments separately revealed an apparent discrepancy: scene-based size expectations affected both sensitivity and criterion in **Experiment 1**, in which the object reappeared with the congruent size on most trials, but they only affected bias in **Experiments 2 & 3**, in which expectations were violated on a substantial proportion of trials. To clarify whether this difference was statistically significant, we ran three separate mixed ANOVAs on the three behavioral measures of interest (accuracy, sensitivity and criterion), with Congruency as within-subject, and Probability/experiment as between-subject factors. Results are shown in **Table 1**: for accuracy, we found a significant main effect of Congruency ($F_{1,148} = 9.42$, $p = 0.003$, $\eta^2_p = 0.060$), and no interaction between Probability and Congruency ($F_{2,148} = 1.81$, $p = 0.167$, $\eta^2_p = 0.024$). This shows that accuracy on the discrimination task was similarly affected by size (in)congruencies in all three experiments (i.e., regardless of short-term probabilities). For sensitivity, likewise, we found a significant main effect of Congruency ($F_{1,148} = 12.57$, $p = 0.001$, $\eta^2_p = 0.078$) and no significant interaction between Probability and Congruency ($F_{2,148} = 1.77$, $p = 0.174$, $\eta^2_p = 0.023$).

Effect on Accuracy	df	F	p	η^2_p
Congruency	1, 148	9.42	**0.003 ****	0.06
Probability	2, 148	1.95	0.145	0.03
Congr. x Prob.	2, 148	1.81	0.167	0.02
Effect on sensitivity	df	F	p	η^2_p
Congruency	1, 148	12.57	**0.001 ****	0.08
Probability	2, 148	1.78	0.172	0.02
Congr. x Prob.	2, 148	1.77	0.174	0.02
Effect on Criterion	df	F	p	η^2_p
Congruency	1, 148	41.73	**< 0.001 *****	0.22
Probability	2, 148	0.86	0.424	0.01
Congr. x Prob.	2, 148	0.48	0.617	0.01

Table 1: Results of the ANOVA including Congruency and Probability. Significant effects are highlighted in boldface. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The effect of size congruency on perceptual sensitivity, then, was not significantly altered by the probability of the object appearing with the congruent size. For criterion, we found a significant main effect of Congruency ($F_{1,148} = 41.73$, $p < 0.001$, $\eta^2_p = 0.220$), but no interaction ($F_{2,148} = 0.48$, $p = 0.617$, $\eta^2_p = 0.006$), confirming the interpretation of the individual experiments by demonstrating that size congruency strongly and consistently influenced participants' criterion across all three experiments.

Overall, this analysis indicates that across the three experiments, the effect of congruency on all three behavioral measures (accuracy, sensitivity and criterion) was consistent. Regardless of the short-term probability with which they were shown during the experiment, then, objects with an incongruent size were still perceived as such, yielding similar effects on task performance.

Final survey data

The purpose of the final survey questions was to assess to what extent participants were aware of the experimental manipulation: how much they paid attention to the sequence of scene viewpoints before the target object appeared, how much they actively tried to predict the final object viewpoint, and their estimate of the probability of the object appearing with the congruent size (see **Methods** for the actual questions asked).

Table S1 reports participants' mean responses for each of the questions, together with their correlations with the difference between Congruent and Incongruent trials (in accuracy, criterion and sensitivity) across subjects. Between-subject Welch ANOVAs revealed that none of the two Likert survey items significantly differed across experiments (Attention to Sequence: $F(2, 98.18) = 0.91$, $p = 0.406$, $\eta^2 = 0.013$; Object Prediction: $F(2, 97.67) = 0.96$, $p = 0.388$, $\eta^2 = 0.012$). This suggests that participants did not adopt a deliberate strategy of paying more or less attention to the scene, or actively trying to predict the object, depending on the probability of the prediction being accurate. Interestingly, their estimates of the probability of the object appearing with the congruent size did not differ across experiments either ($F(2, 98.11) = 0.26$, $p = 0.774$, $\eta^2 = 0.003$). Thus, we found no evidence that participants were tracking how often the contextual expectation was respected or violated, despite this expectation's impact on their responses.

Moreover, we found no evidence for correlations between the responses to any of the survey questions and the behavioral differences in accuracy, criterion, or sensitivity (**Table S1**). In fact, the only correlations between a survey question and behavioral effect that were marginally significant (both in Exp. 1: object prediction and Δ accuracy, $r = -0.29$, $p = 0.041$, and object prediction and Δ sensitivity, $r = -0.33$, $p = 0.021$, uncorrected for multiple comparisons) were *negative* correlations. The behavioral effects we found, then, did not seem to depend on participants' awareness of the experimental manipulation. Possible differences in their strategies across experiments were then likely automatic, and not the product of conscious deliberation.

Discussion

The retinal sizes of objects in the real world depend on the distance from which they are viewed. Scene context provides a reference frame to estimate that distance, and changes in the reference frame (e.g., while moving) thus have the potential to guide our predictions of object size transformations. In the present work, we found that participants responded differently in an orthogonal perceptual task, depending on whether an object appeared in a size that was expected or unexpected given the scene context. This difference in responses demonstrates that they perceived the (in)congruency of the object size, which could only be derived from the shifting viewpoint of the scene. Scene context, then, drove the automatic rescaling of object representations.

Further evidence for the automaticity of scene-driven rescaling is the fact that the behavioral effects were largely consistent across experiments, even when scene-driven expectations were violated on a large proportion of trials. This suggests that the rescaling of object representations was primarily driven by long-term expectations derived from regularities of the real world, which overruled short-term regularities observed during the experiment. Additionally, we found that the effect of congruency on task performance (accuracy, sensitivity, or criterion) did not correlate with participants' reports of paying attention to the contextual sequence or explicitly trying to predict the object view, nor to their awareness of the frequency of congruent trials (**Table S1**). The effect of scene context on object representations, then, appears to be independent of participants' conscious behavioral strategy.

Interestingly, the effect of scene-driven size expectations in our task reliably resulted in a shift of criterion on trials in which the object did not match those expectations. While participants were overall slightly biased towards reporting that the two probe views were the same (possibly due to the generally small differences between them), this bias was reduced on Incongruent trials. That is, they responded 'different' more often. A possible explanation for this criterion shift is that the size incongruency between the initial and final view of the object was perceived as a change, drawing their responses towards 'different'. Importantly, however, participants were robustly above chance in all experiments, as this was a precondition for inclusion in the analysis. This indicates that they had not misunderstood the task, and were not actively trying to predict the upcoming object size, which would have resulted in chance performance. Predictions of object size, instead, seem to have occurred automatically, involuntarily influencing their responses.

Expectations also affected perceptual sensitivity, with a reduced sensitivity on Incongruent trials. This effect was particularly reliable in Experiment 1 (75% probability), although the interaction between probability and congruency was not significant. It is thus possible that the effect of congruency on perceptual sensitivity varies according to the task context. One perspective is to regard the response interference caused by incongruent object sizes as relating to surprise. When incongruent object sizes were rare, as in Experiment 1, incongruent objects would be more surprising, potentially reducing participants' focus on the stimulus features that were relevant for performing the main perceptual discrimination task. Interestingly, criterion did not show a similar variation across experiments in the present study. In contrast, our previous results on mental rotation (Aldegheri et al., 2023) showed a significant interaction between congruency and the probability of congruent object views for both criterion and sensitivity. The reason for this discrepancy is not clear and may be the result of idiosyncrasies of the stimuli and tasks used in the two studies. While both studies are consistent in showing the existence and automaticity of object representation transformations induced by scene context, then, the specific ways in which these transformations affect task performance might depend on specific features of the task.

The present results support the idea that transformations of a scene context can induce coherent transformations of the objects within that scene. This applies to both rotation and translation, two rigid spatial transformations that commonly occur in the real world. This is in contrast to *non-rigid* transformations (that alter the shapes of objects), such as squeezing, breaking, or melting. Numerous studies have shown that humans are able to predict the consequences of such transformations (Kourtzi & Shiffrar, 2001; Hahn et al., 2009; Spröte &

Fleming, 2016; Spröte et al., 2016; Hafri et al., 2022), suggesting that we can simulate them in our minds similarly to rigid transformations. In the real world, however, the dependency between objects and their context is less clear for non-rigid transformations. Transformations like breaking into pieces, or being bent out of shape, tend to affect single objects in a scene, and they are usually caused by agents or other objects rather than by a global transformation of the entire scene. The representations of interactions between objects that result in rigid and non-rigid transformations, then, might be qualitatively different. For example, rigid transformations can be represented hierarchically, with the transformation of a scene affecting its objects, and the objects' transformation affecting their parts in turn (Hinton & Parsons, 1981; Hinton, 1990; Gklezakos & Rao, 2022; Hinton, 2023; Fisher & Rao, 2023; Shewmake et al., 2023). Such a hierarchical representation might be less suited for non-rigid transformations, in which the dependencies between objects and their parts are more complex. When an object breaks, for example, not all of its parts are affected equally, and the way they are affected largely depends on the interaction that caused the breaking (e.g. the object falling on the floor, or being hit with a hammer). It is possible, then, that rigid transformations affecting a scene and its parts are represented hierarchically, while interactions such as breaking or squeezing are represented in a 'flat' manner, with individual objects in a scene influencing each other. Prior work has found evidence for such representations of physical or social interactions between objects or agents (Hafri & Firestone, 2021; Little & Firestone, 2021; Little & Gureckis, 2023; Malik & Isik, 2023). Future research should clarify the differences in how these disparate interactions are represented in human visual perception, to elucidate the nature of our internal models of the physical world.

The present results provide additional evidence that predictions of spatial transformations can be driven by contextual information, rather than purely internal operations. In our previous studies (Aldegheri et al., 2023, 2025), we found that mental rotation, a widely studied mental transformation, could be driven by the changing viewpoint of a scene. Here, we find that this generalizes to mental translation. These findings raise the question of whether the context-driven transformations that we observed involve similar representations as those implicated in more classic purely-internal transformations of isolated objects. Voluntary mental transformations of isolated objects are believed to involve the manipulation of image-like representations (Cooper & Shepard, 1973; Shepard & Cooper, 1982; Koriat & Norman, 1984, 1988; Stewart et al., 2022). One possibility is that the context-driven transformations we observed in our studies also involve the creation of a 'mental image' of the object at the congruent orientation or size. An alternative possibility is that, instead, what was transformed along with the scene was a spatial *reference frame*. According to this account, the changing scene context establishes a set of spatial coordinates along which the position and orientation of objects can be represented. Graf (2006) reviews a series of findings indicating that human observers can mentally transform spatial coordinate frames, facilitating the perception of objects which are oriented, translated or scaled consistently with them. For example, recognition of objects in a specific orientation (Graf et al., 2005) or size (Larsen & Bundesen, 1978) is facilitated after the observer has seen other objects with the same orientation or size. Contextual information, then, can help establish spatial reference frames which generalize across objects. Crucially, these reference frames can also be determined by scene context (Hinton & Parsons, 1988; Humphrey & Jolicoeur, 1993; Christou et al., 2003), suggesting that scene transformations in our study might be influencing object representations through a similar mechanism. If this is the case, then what is being updated behind the occluder is not a mental image of the entire object, but an abstracted representation of its scale or orientation. This would be consistent with results from visual object tracking under occlusion, which show that only the location of the object, and not other features, are tracked behind the occluder (Scholl & Pylyshyn, 1999; Teichmann et al., 2022). A possible way to distinguish between an image-like and a purely spatial representation would be to determine whether the behavioral effect of size or orientation (in)congruency generalizes across objects. If a different object (e.g., a bed instead of a couch) reappeared after the occlusion, would it make any difference to performance in an orthogonal task if its size or orientation were (in)congruent? If only a spatial reference frame was being updated with the scene, we would expect the behavioral effect of spatial (in)congruencies to generalize across objects, as observed in previous studies (Graf et al., 2005; Larsen & Bundesen, 1978). If, on the other hand, subjects were updating a

picture-like representation of the object behind the occluder, a different object would be perceived as 'incongruent' regardless of its size or orientation. Further research is needed to distinguish between these two hypotheses.

In conclusion, the present results show that scene viewpoint information can automatically drive predictions of object size - or translation in depth. These findings generalize our previous results using rotation, thereby suggesting that scene context drives dynamic predictions of object appearance across a range of transformations. These predictions affect responses in an orthogonal behavioral task, even when they violate short-term expectations, pointing toward the automaticity of such predictions based on real-world regularities. Scene context, then, might play a general role in providing a reference frame for different mental transformations of objects. Given the highly structured nature of our everyday environments, this might be an important mechanism supporting our interaction with objects in naturalistic vision.

Funding

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 725970) awarded to Marius Peelen.

Data availability

All data and stimuli are publicly available at https://osf.io/dzf6k/. Code for running the online experiments and analyzing the data is publicly available at https://github.com/GAldegheri/scenecontext-zoom.

References

Aldegheri, G., Gayet, S., & Peelen, M. V. (2023). Scene context automatically drives predictions of object transformations. *Cognition*, *238*, 105521.

Aldegheri, G., Gayet, S., & Peelen, M. V. (2025). *Dynamic context-based updating of object representations in visual cortex*. bioRxiv.

Bennett, D. (2002). Evidence for a pre-match 'mental translation' on a form-matching task. *Journal of Vision*, *2*(7), 50. https://doi.org/10.1167/2.7.50

Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 214–220. https://doi.org/10.1037/0096-1523.1.3.214

Bundesen, C., Larsen, A., & Farrell, J. E. (1983). Visual Apparent Movement: Transformations of Size and Orientation. *Perception*, *12*(5), 549–558. https://doi.org/10.1068/p120549

Christou, C. G., Tjan, B. S., & Bülthoff, H. H. (2003). Extrinsic cues aid shape recognition from novel viewpoints. *Journal of Vision*, *3*(3), 1. https://doi.org/10.1167/3.3.1

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual Information Processing* (pp. 75–176). Academic Press. https://doi.org/10.1016/B978-0-12-170150-5.50009-3

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12.

Fisher, A., & Rao, R. P. (2023). Recursive neural programs: A differentiable framework for learning compositional part-whole hierarchies and image grammars. *PNAS Nexus*, *2*(11), pgad337.

Gayet, S., Battistoni, E., Thorat, S., & Peelen, M. V. (2024). Searching near and far: The attentional template incorporates viewing distance. *Journal of Experimental Psychology: Human Perception and Performance*, *50*(2), 216–231. https://doi.org/10.1037/xhp0001172

Gayet, S., & Peelen, M. V. (2022). Preparatory attention incorporates contextual expectations. *Current Biology*, *32*(3), 687–692.

Gklezakos, D. C., & Rao, R. P. (2022). Active Predictive Coding Networks: A Neural Solution to the Problem of Learning Reference Frames and Part-Whole Hierarchies. *arXiv Preprint arXiv:2201.08813*.

Graf, M. (2006). Coordinate transformations in object recognition. *Psychological Bulletin*, *132*, 920–945. https://doi.org/10.1037/0033-2909.132.6.920

Graf, M., Kaping, D., & Bülthoff, H. H. (2005). Orientation congruency effects for familiar objects: Coordinate transformations in object recognition. *Psychological Science*, *16*(3), 214–221. https://doi.org/10.1111/j.0956-7976.2005.00806.x

Hafri, A., Boger, T., & Firestone, C. (2022). Melting ice with your mind: Representational momentum for physical states. *Psychological Science*, *33*(5), 725–735.

Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*, *25*(6), 475–492.

Hahn, U., Close, J., & Graf, M. (2009). Transformation direction influences shape-similarity judgments. *Psychological Science*, *20*(4), 447–454. https://doi.org/10.1111/j.1467-9280.2009.02310.x

Hamrick, J. B., & Griffiths, T. (2014). What to simulate? Inferring the right direction for mental rotation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*(36). https://escholarship.org/uc/item/064367d4

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., & Smith, N. J. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51.

Hinton, G. (2023). How to represent part-whole hierarchies in a neural network. *Neural Computation*, *35*(3), 413–452.

Hinton, G. E. (1990). Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence*, *46*(1–2), 47–75.

Hinton, G. E., & Parsons, L. M. (1981). Frames of reference and mental imagery. In *Attention and performance IX* (pp. 261–277).

Hinton, G. E., & Parsons, L. M. (1988). Scene-based and viewer-centered representations for comparing shapes. *Cognition*, *30*(1), 1–35.

Humphrey, G. K., & Jolicoeur, P. (1993). An examination of the effects of axis foreshortening, monocular depth cues, and visual field on object identification. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *46A*, 137–159. https://doi.org/10.1080/14640749308401070

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(03), 90–95.

Koriat, A., & Norman, J. (1984). What is rotated in mental rotation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 421–434. https://doi.org/10.1037/0278-7393.10.3.421

Koriat, A., & Norman, J. (1988). Frames and images: Sequential effects in mental rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 93–111. https://doi.org/10.1037/0278-7393.14.1.93

Kourtzi, Z., & Shiffrar, M. (2001). Visual representation of malleable and rigid objects that deform as they rotate. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 335–355. https://doi.org/10.1037/0096-1523.27.2.335

Kuroki, D. (2021). A new jsPsych plugin for psychophysics, providing accurate display duration and stimulus onset asynchrony. *Behavior Research Methods*, *53*(1), 301–310.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*(3), 389–412.

Larsen, A. (2014). Deconstructing mental rotation. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1072–1091. https://doi.org/10.1037/a0035648

Larsen, A., & Bundesen, C. (1978). Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 1–20. https://doi.org/10.1037/0096-1523.4.1.1

Larsen, A., & Bundesen, C. (1998). Effects of spatial separation in visual pattern matching: Evidence on the role of mental translation. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 719–731. https://doi.org/10.1037/0096-1523.24.3.719

Larsen, A., & Bundesen, C. (2009). Common mechanisms in apparent motion perception and visual pattern matching. *Scandinavian Journal of Psychology*, *50*(6), 526–534. https://doi.org/10.1111/j.1467-9450.2009.00782.x

Leibowitz, H., Brislin, R., Perlmutrer, L., & Hennessy, R. (1969). Ponzo perspective illusion as a manifestation of space perception. *Science*, *166*(3909), 1174–1176.

Little, P. C., & Firestone, C. (2021). Physically implied surfaces. *Psychological Science*, *32*(5), 799–808.

Little, P. C., & Gureckis, T. M. (2023). *Inferring the existence of objects from their physical interactions*. https://par.nsf.gov/biblio/10466904

Malik, M., & Isik, L. (2023). Relational visual representations underlie human social interaction recognition. *Nature Communications*, *14*(1), 7317. https://doi.org/10.1038/s41467-023-43156-8

McKinney, W. (2011). pandas: A foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, *14*(9), 1–9.

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.

Schmidt, F., Spröte, P., & Fleming, R. W. (2016). Perception of shape and space across rigid transformations. *Vision Research*, *126*, 318–329. https://doi.org/10.1016/j.visres.2015.04.011

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking Multiple Items Through Occlusion: Clues to Visual Objecthood. *Cognitive Psychology*, *38*(2), 259–290. https://doi.org/10.1006/cogp.1998.0698

Sekuler, R., & Nash, D. (1972). Speed of size scaling in human vision. *Psychonomic Science*, *27*, 93–94. https://doi.org/10.3758/BF03328898

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations* (pp. viii, 364). The MIT Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Shewmake, C. A., Buracas, D., Lillemark, H., Shin, J., Bekkers, E. J., Miolane, N., & Olshausen, B. (2023). Visual Scene Representation with Hierarchical Equivariant Sparse Coding. *NeurIPS 2023 Workshop on Symmetry and Geometry in Neural Representations*. https://openreview.net/forum?id=TF2RcrcTP2

Spröte, P., & Fleming, R. W. (2016). Bent out of shape: The visual inference of non-rigid shape transformations applied to objects. *Vision Research*, *126*, 330–346.

Spröte, P., Schmidt, F., & Fleming, R. W. (2016). Visual perception of shape altered by inferred causal history. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep36245

Stewart, E. E. M., Hartmann, F. T., Morgenstern, Y., Storrs, K. R., Maiello, G., & Fleming, R. W. (2022). Mental object rotation based on two-dimensional visual representations. *Current Biology*, *32*(21), R1224–R1225. https://doi.org/10.1016/j.cub.2022.09.036

Teichmann, L., Moerel, D., Rich, A. N., & Baker, C. I. (2022). The nature of neural object representations during dynamic occlusion. *Cortex*, *153*, 66–86.

Vallat, R. (2018). Pingouin: Statistics in Python. *Journal of Open Source Software*, *3*(31), 1026.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272.

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021.

Wetherill, G., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology*, *18*(1), 1–10.

Xue, J., Li, C., Quan, C., Lu, Y., Yue, J., & Zhang, C. (2017). Uncovering the cognitive processes underlying mental rotation: An eye-movement study. *Scientific Reports*, *7*(1), 1–12.

Yeh, L.-C., Gayet, S., Kaiser, D., & Peelen, M. V. (2024). The neural time course of size constancy in natural scenes. *bioRxiv*, 2024–09.

Yildiz, G. Y., Sperandio, I., Kettle, C., & Chouinard, P. A. (2021). A review on various explanations of Ponzo-like illusions. *Psychonomic Bulletin & Review*, 1–28.

Supplementary Materials

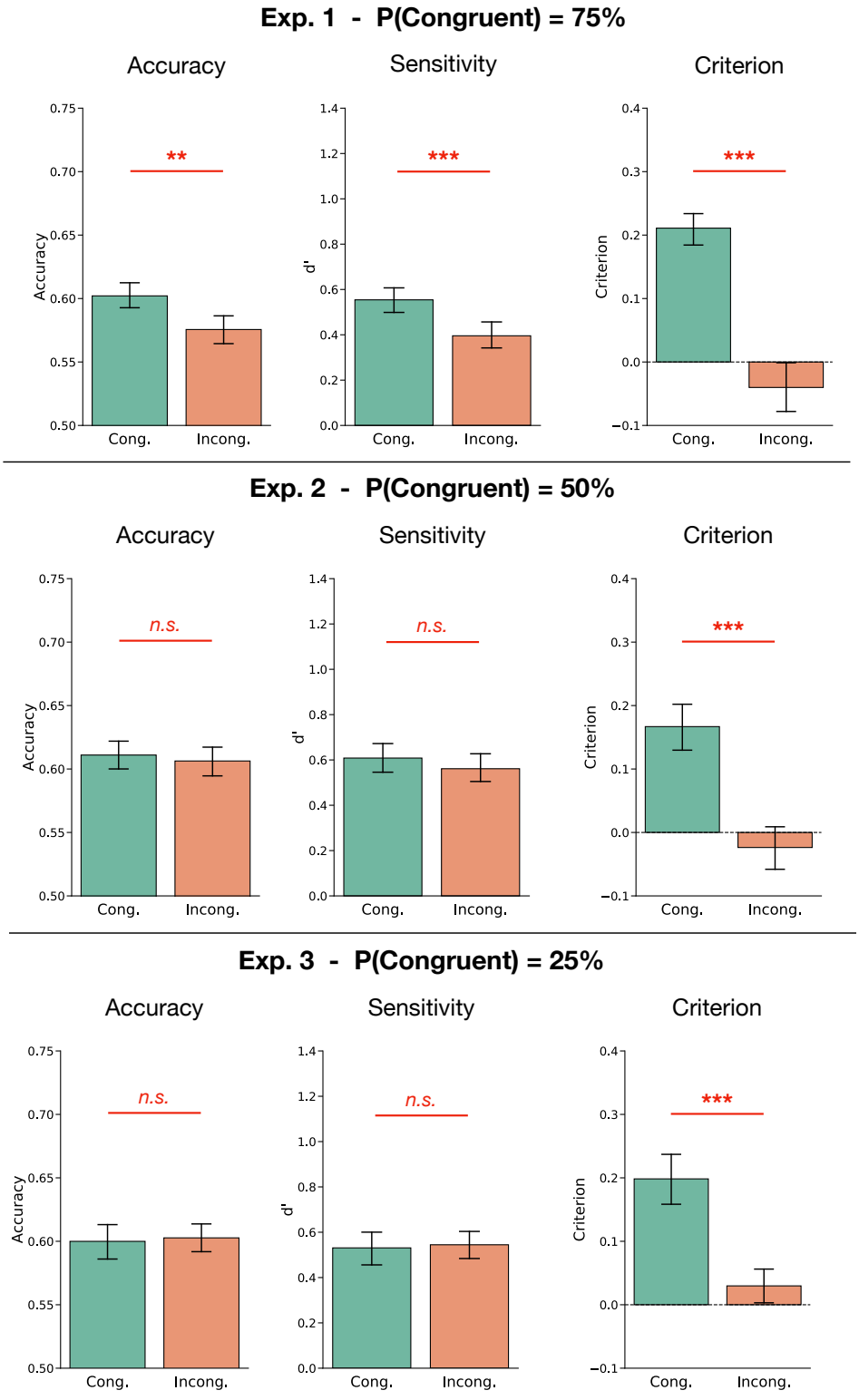

Figure S1: Results of the three experiments without any participant exclusions based on performance. For each experiment, mean (and SEM) accuracy, d' and criterion are reported for the Congruent and Incongruent conditions. ** $p < 0.01$, *** $p < 0.001$.

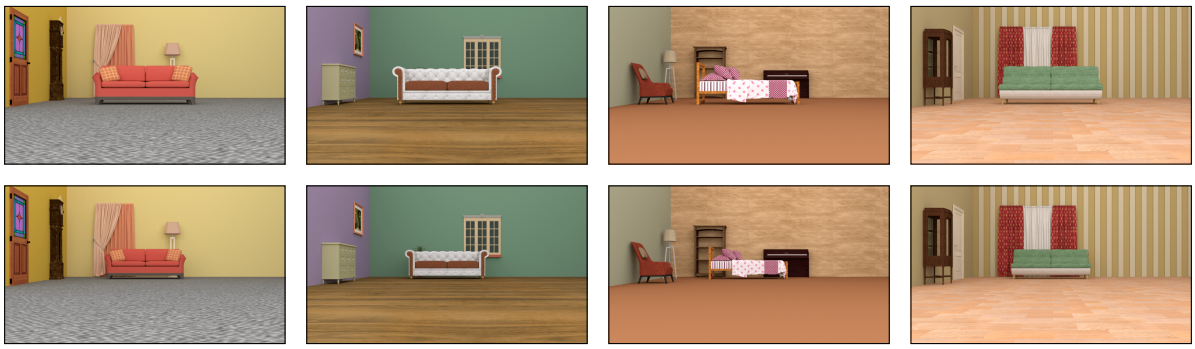

Figure S2: The four different scene exemplars used in the study. Each scene is shown with the object in the ‘near’ (top) and ‘far’ (bottom) initial position. See main text for details.

	Experiment 1 **P(Congruent) = 75%**	**Experiment 2** **P(Congruent) = 50%**	**Experiment 3** **P(Congruent) = 25%**
Attention to Sequence 1-7 Likert scale	4.42 ± 0.19	4.30 ± 0.21	4.02 ± 0.23
Correlation with Δ accuracy	r = 0.17, p = 0.23, BF_{01} = 2.79	r = 0.03, p = 0.83, BF_{01} = 5.55	r = -0.09, p = 0.52, BF_{01} = 4.67
Correlation with Δ sensitivity	r = 0.13, p = 0.38, BF_{01} = 3.89	r = 0.05, p = 0.71, BF_{01} = 5.29	r = -0.11, p = 0.46, BF_{01} = 4.39
Correlation with Δ criterion	r = 0.07, p = 0.62, BF_{01} = 5.05	r = 0.04, p = 0.75, BF_{01} = 5.40	r = 0.05, p = 0.74, BF_{01} = 5.43
Object Prediction 1-7 Likert scale	3.53 ± 0.24	3.76 ± 0.22	3.31 ± 0.24
Correlation with Δ accuracy	**r = -0.29, p = 0.04, BF_{01} = 0.74 ***	r = -0.06, p = 0.67, BF_{01} = 5.21	r = 0.09, p = 0.53, BF_{01} = 4.74
Correlation with Δ sensitivity	**r = -0.33, p = 0.02, BF_{01} = 0.42 ***	r = 0.02, p = 0.86, BF_{01} = 5.59	r = 0.10, p = 0.48, BF_{01} = 4.50
Correlation with Δ criterion	r = 0.04, p = 0.79, BF_{01} = 5.40	r = 0.17, p = 0.23, BF_{01} = 2.85	r = -0.19, p = 0.19, BF_{01} = 2.50
Probability Estimate Percentage	56.96 ± 2.45	54.50 ± 2.39	55.84 ± 2.94
Correlation with Δ accuracy	r = 0.03, p = 0.83, BF_{01} = 5.55	r = -0.11, p = 0.43, BF_{01} = 4.18	r = -0.02, p = 0.91, BF_{01} = 5.68
Correlation with Δ sensitivity	r = -0.03, p = 0.83, BF_{01} = 5.55	r = -0.13, p = 0.37, BF_{01} = 3.86	r = 0.04, p = 0.76, BF_{01} = 5.46
Correlation with Δ criterion	r = -0.06, p = 0.69, BF_{01} = 5.26	r = 0.10, p = 0.48, BF_{01} = 4.44	r = -0.12, p = 0.38, BF_{01} = 3.95

Table S1: Mean responses (and SEM) to our final survey questions, and Pearson's r correlation with the behavioral effects (Congruent – Incongruent trials) for both criterion and d'. * $p < 0.05$ (uncorrected for multiple comparisons).