Cognition 238 (2023) 105521

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Scene context automatically drives predictions of object transformations

Giacomo Aldegheri[a,b,*], Surya Gayet[a,c], Marius V. Peelen[a]

[a] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Thomas van Aquinostraat 4, Nijmegen 6525 GD, the Netherlands
[b] Department of Psychology, Amsterdam Brain & Cognition Center, University of Amsterdam, Nieuwe Achtergracht 129-B, Amsterdam 1018 WS, the Netherlands
[c] Department of Experimental Psychology, Helmholtz Institute, Utrecht University, Heidelberglaan 1, Utrecht 3584 CS, the Netherlands

ARTICLE INFO

Keywords:
Scene perception
Mental rotation
Visual expectations
Object perception

ABSTRACT

As our viewpoint changes, the whole scene around us rotates coherently. This allows us to predict how one part of a scene (e.g., an object) will change by observing other parts (e.g., the scene background). While human object perception is known to be strongly context-dependent, previous research has largely focused on how scene context can disambiguate fixed object properties, such as identity (e.g., a car is easier to recognize on a road than on a beach). It remains an open question whether object representations are updated dynamically based on the surrounding scene context, for example across changes in viewpoint. Here, we tested whether human observers dynamically and automatically predict the appearance of objects based on the orientation of the background scene. In three behavioral experiments ($N = 152$), we temporarily occluded objects within scenes that rotated. Upon the objects' reappearance, participants had to perform a perceptual discrimination task, which did not require taking the scene rotation into account. Performance on this orthogonal task strongly depended on whether objects reappeared rotated coherently with the surrounding scene or not. This effect persisted even when a majority of trials violated this real-world contingency between scene and object, showcasing the automaticity of these scene-based predictions. These findings indicate that contextual information plays an important role in predicting object transformations in structured real-world environments.

1. Introduction

The retinal image of objects in our everyday environments changes as we move. Predicting such changes is crucial for acting quickly and efficiently within a dynamic world. For this purpose, we are equipped with internal representations of objects that can be transformed similarly to those objects (Higgins, Racanière, & Rezende, 2022; Shepard, 1984, 2001). For example, we can mentally rotate (Shepard & Metzler, 1971), translate (Larsen & Bundesen, 1998) or rescale object representations (Bundesen & Larsen, 1975) to reliably predict how objects will look from novel viewpoints.

In the real world, object transformations such as changes in orientation or position are strongly constrained by scene context. For example, as we navigate an environment, the perceived orientation of stationary objects will change jointly with the orientation of the scene (e.g., the orientations of walls). Real-world vision is thus highly redundant: by observing one part of the scene, it is possible to predict another. While previous research has investigated how static scene context informs predictions about object identity (e.g., a car will be better recognized on a road than on the sea; Bar, 2004; Oliva & Torralba, 2007; Brandman & Peelen, 2017; Võ, Boettcher, & Draschkow, 2019), not much is known about how scene context dynamically guides predictions about spatial transformations of objects. Here, we test the hypothesis that observers automatically predict object orientation based on scene context, as measured by improved performance in an orthogonal perceptual discrimination task.

Exploiting scene context to predict object transformations could alleviate the computational burden of mentally transforming objects, which is a slow and effortful process (Just & Carpenter, 1976; Larsen, 2014; Xue et al., 2017). For example, the amount of rotation of the scene would directly determine the rotation of the object. This stands in stark contrast with the requirements of traditional mental rotation paradigms, in which participants need to determine the angle and direction of rotation that is required to solve the task (Hamrick & Griffiths, 2014). Rather than imagining object transformations on a 'mental screen' separated from current visual input, such context-driven transformations would amount to a task of *completing* partial information present in a scene. This process could be particularly beneficial in real-

* Corresponding author at: Department of Psychology, Amsterdam Brain & Cognition Center, University of Amsterdam, Nieuwe Achtergracht 129-B, Amsterdam 1018 WS, the Netherlands.
E-mail address: giacomo.aldegheri@gmail.com (G. Aldegheri).

https://doi.org/10.1016/j.cognition.2023.105521
Received 18 April 2023; Received in revised form 7 June 2023; Accepted 8 June 2023
Available online 22 June 2023

world situations where internal representations and external context are integrated, for instance when keeping track of temporarily occluded objects (Munton, 2022; Scholl & Pylyshyn, 1999).

Here, in three behavioral experiments, we investigated whether human participants automatically predict the appearance of a rotated object, based exclusively on the orientation of the surrounding scene. We designed an experimental paradigm in which an object (a bed or couch) was shown in the context of a realistic indoor scene (Fig. 1). While the viewpoint on the scene kept changing, the object was temporarily hidden by an occluder, and then reappeared. When it reappeared, the object could either be oriented consistently (*Congruent* condition) or inconsistently (*Incongruent* condition) with the scene's new viewpoint. Importantly, the exact same view of the object could be congruent or incongruent (on a given trial) depending on (A) the object's starting position and (B) the amount of scene rotation, thus ruling out any stimulus-related differences between conditions. Moreover, because the angle of scene rotation was varied across trials, participants could not predict the object's congruent orientation by simply extrapolating the rotation of the object alone. Therefore, any difference in processing between congruent and incongruent objects implies that observers inferred the appearance of the object from changes in the surrounding scene.

In order to assess how scene-object orientation congruency influenced object processing, we measured participants' performance on a perceptual discrimination task (Fig. 1B). Upon reappearance, the object was displayed twice, very briefly, and participants had to indicate whether or not the object had the exact same orientation in these two displays (same/different task). We compared accuracy in this task between congruent and incongruent trials. Importantly, this task was fully orthogonal to the congruency manipulation, and participants were not explicitly instructed to exploit scene information, nor were they incentivized to predict the upcoming object view. We hypothesized that participants would be more accurate in the congruent than the incongruent condition, reflecting the facilitatory influence of scene-based predictions (De Lange, Heilbron, & Kok, 2018).

To summarize our results, we found that perceptual discrimination accuracy was higher when the object reappeared in a congruent as compared to an incongruent orientation relative to the scene, providing evidence that participants used scene orientation to predict object orientation. In Experiment 1, this congruency effect occurred when the congruent orientation appeared on a majority of trials (75%). To disentangle the role of real-world regularities and short-term experiment contingencies in driving the effect, in subsequent experiments we presented congruent and incongruent orientations with equal probability (50%; Experiment 2), and even reversed the real-world regularity by showing the congruent orientation on a minority of trials (25%; Experiment 3). Despite manipulating these short-term contingencies, the congruency effect still followed real-world (rather than short-term) regularities, indicating that scene viewpoint influences mental object transformations in an automatic fashion. Moreover, this influence is not abolished or reversed when expectations are violated frequently during the experiment, suggesting that the scene-based predictions derive primarily from knowledge of long-term real-world regularities, and cannot be easily overruled. Automatically predicting object transformations

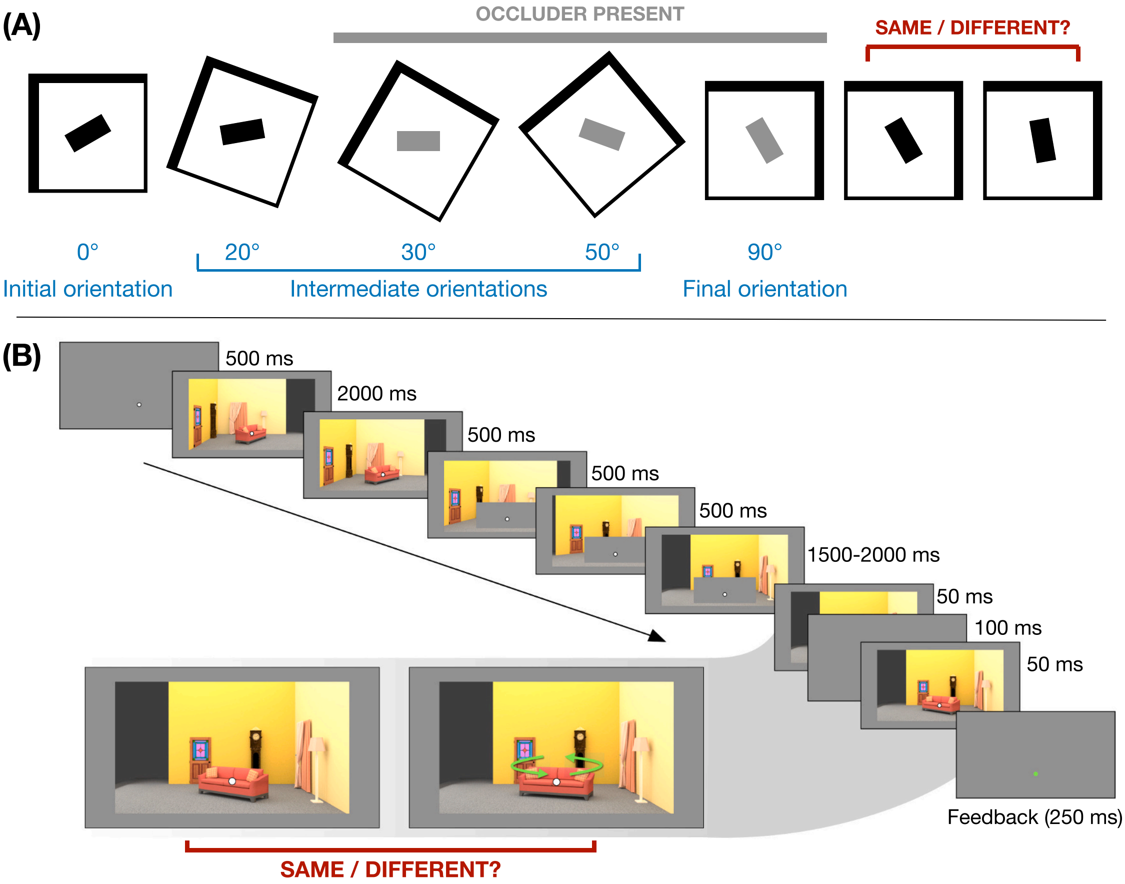

Fig. 1. (A) Schematic of the sequence of scene orientations in an example trial, seen from above. (B) The stimuli shown in in this same example trial, where the total rotation of the scene is "Large" (90°), the orientation of the object upon reappearance is "Congruent" relative to the scene, and the two probes are "Different": the second probe is slightly rotated relative to the first (green arrows added for illustration). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

based on scene context helps to overcome the complexity of the real world by exploiting its regularities.

2. Methods

2.1. 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).

Online participants were recruited on Prolific (Palan & Schitter, 2018), and had to satisfy the following criteria: reside in Europe or the UK, to ensure their time zone was the same as ours and they were participating during day hours; be between 18 and 35 years old; have normal or corrected-to-normal vision; 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 hit rate in our same/different task with 50% was significant (at $\alpha = 0.05$), meaning that they were performing better than chance. We continued data collection until the number of included participants reached 50 for each experiment (in batches of 10–20 participants: in Experiment 3, this led to a final sample size of 52, as more participants in the final batch met the inclusion criteria than expected). Previous pilot studies in the laboratory revealed a large effect size (> 0.8), such that a sample of 23 participants would be sufficient for 95% power. Given that conducting the experiment online might have reduced the effect size, we chose to aim for a larger sample size of 50. In Experiment 1, we excluded 30 participants. Of the included 50 participants, 25 were female, and mean age (and standard deviation) was 26.7 ± 5.1. In Experiment 2, we excluded 33 participants. Of the included 50 participants, 20 were female, and mean age was 24.5 ± 4.3. In Experiment 3, we excluded 56 participants. Of the included 52 participants, 25 were female, 26 male and one participant's demographic information was missing. Mean age was 24.7 ± 4.4.

The high exclusion rate was likely due to several reasons: primarily, we limited the maximum possible orientation difference between the two probes to 20°, to avoid exceeding 1/3 of the difference between congruent and incongruent views (60°). This meant that the staircase was limited in its ability to adjust to participants with a higher discrimination threshold. Moreover, we kept a very short presentation time (50 ms) for the two probes, in order to reduce the influence of deliberate judgment and find evidence of a perceptual representation of the object's updated appearance, thereby making the task more challenging. Importantly, however, all the results reported here held true with no participant exclusions (see Supplementary Materials S.2, **Fig. S3**).

2.2. Stimuli

The stimuli were based on 8 different indoor scenes (**Fig. S1**) 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 contained various other objects, adjacent to the walls, and different textures on the walls and floor. The central object was a couch in half of the scenes, and a bed in the other half. The size of the central object was the same across scenes. For each scene, a sequence of different viewpoints was rendered, by rotating the scene around the vertical axis between 0° to 90° in steps of 5°. The two walls were oriented such that the scene was fully visible from all these viewpoints. All scene images were presented at a resolution of 960 × 540 pixels.

2.3. Procedure

Each trial (Fig. 1) began with a fixation dot (which was always present during the trial, radius 5 pixels) for 500 ms, followed by the first orientation of the scene for 2000 ms, the 3 intermediate orientations 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 (see Fig. 1A for a schematic representation). The occluder had the height and width of the largest possible view of the object, plus a margin (horizontal: 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 orientation.

After the final orientation of the scene was shown, with the object still fully occluded, the object was briefly flashed twice (within the scene) for 50 ms, with a 100 ms inter-stimulus interval in between. We refer to these two brief presentations of the object as the *probes*. The task of the participants 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 answer 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 views of the objects exclusively, but that they should also pay attention to the preceding sequence of images, to ensure they would not completely disengage during the seconds preceding the probes.

The orientation of the first probe object was randomly sampled from a normal distribution centered around the Congruent or Incongruent object viewpoint, with a standard deviation of 1°, to add a small amount of jitter, and then rounded to the nearest integer. The second probe was exactly the same as the first probe in half of the trials ('same' trials). In the other half of the trials ('different' trials), the second probe was rotated around the vertical axis relative to the first (see Fig. 1B, bottom left), clockwise or counterclockwise with equal probability.

The orientation difference on the '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 Congruency conditions to ensure average performance around 70% correct (Wetherill & Levitt, 1965) across conditions, while still allowing for accuracy differences between the Congruent and Incongruent conditions. The motivation for using a single staircase across conditions, rather than two separate interleaved staircases on Congruent and Incongruent trials as more commonly done (e.g. Kok, Jehee, & De Lange, 2012), was to keep the physical stimuli exactly the same in both conditions, thus ensuring the full orthogonality of the Congruency manipulation from the physically presented stimuli. Allowing the probe orientation differences probes to vary between conditions could have drawn participants' attention to the manipulation, possibly leading to changes in response strategies. Stimulus intensity (orientation difference between probes) was adjusted after both 'same' and 'different' trials. The starting value for the staircase was 10°, step size was 1° (lowered to 0.5° after 3 staircase reversals) and the minimum and maximum possible orientation differences shown were 0.5° and 20°, respectively. The means and standard deviations of the angle differences reached by the staircase in the second half of trials, in each of the three experiments, were 12.76° ± 4.64, 11.96° ± 5.18, and 14.11° ± 4.90 respectively.

Each experiment lasted about 30 min in total, divided in 8 blocks, and participants were encouraged to take a short break after the end of each block. Before the experiment began, participants read 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 50 ms, which is the presentation time used in the main experiment. This allowed participants to familiarize with the task with an initially less challenging presentation time.

2.4. Experimental design

Trials varied along three different factors (Fig. 2): Congruency (Congruent, Incongruent), initial Object Orientation relative to the scene (6 angles: 0°, 60°, 120°, 180°, 240°, 300°), Scene Rotation (Small, Large) and Scene (1 of 4 different scene exemplars, one of two subsets of the 8 total views, selected randomly for each participant).

The overall proportion of Congruent and Incongruent trials varied depending on the experiment (75% of total trials in Exp. 1, 50% in Exp. 2, and 25% in Exp. 3). All the other factors were fully balanced within the Congruent and Incongruent trials, meaning that each of four partitions of the trials (variably assigned to either Congruent or Incongruent depending on the experiment) were equally divided among each combination of Object Orientation, Scene Rotation and Scene (6 × 2 × 4 = 48 trials for each partition, resulting in 192 trials in total). All these trials were presented in random order throughout the experiment.

The Incongruent view corresponded, on Small scene rotation trials, to an object that was rotated 60° more than expected, and on Large scene rotation trials, to an object rotated 60° less than expected. The 6 initial object orientations were chosen to be 60° apart, so that the Incongruent view for one orientation corresponded to the Congruent one for another (see also **Fig. S2**). Consequently, the exact same images were presented as Congruent in the context of one trial, and Incongruent in the context of another trial, avoiding any possible confounds due to physical differences between conditions (Fig. 2).

2.5. Post-experiment survey

After completing the experiment, participants were asked three questions about their awareness of the congruency 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 rotated, 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%.

2.6. 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). For the three-way mixed ANOVA in the scene alignment analysis (see **Role of alignment with the scene** in **Results**), we used the R package bruceR (https://CRAN.R-project.org/package=bruceR).

3. Results

3.1. Experiment 1: 75% probability

In the first experiment, participants had a 75% probability of seeing the object reappear in the congruent view. Across conditions, their mean accuracy (and SEM) was 0.68 ± 0.01, indicating that they were able to perform the task, and that the staircase successfully reached the desired accuracy range, around 70%.

In our central analysis, we compared accuracy between Congruent

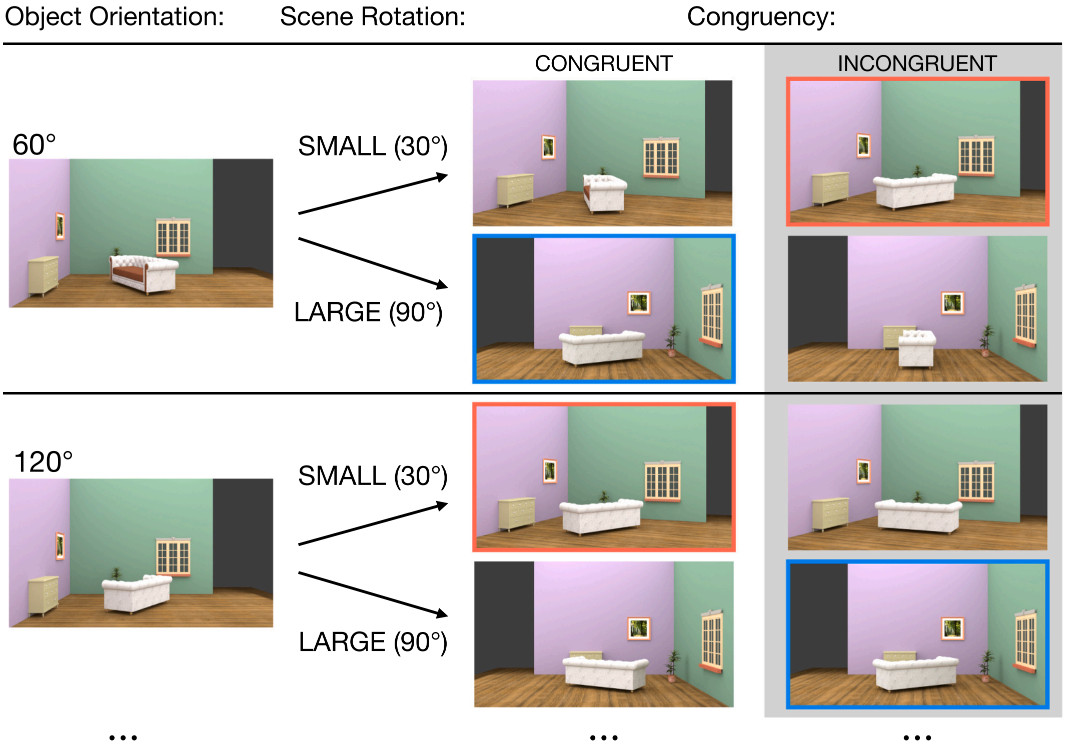

Fig. 2. Illustration of the experimental design, showing the initial orientation of the object relative to the scene, and the final images (after the whole sequence of rotations, including the occlusion period) resulting from a Small or Large rotation on Congruent or Incongruent trials. The images highlighted by the colored frames are examples of the same images appearing as either Congruent or Incongruent on different trials. Note that these are only two out of six possible starting orientations (see **Fig. S2**); throughout the experiment every final image occurred in the Congruent as well as the Incongruent condition.

and Incongruent trials. Participants were significantly more accurate on Congruent than Incongruent trials

(means: 0.71 vs. 0.65; $t_{49} = 5.87$, $p < 0.001$, d = 0.95, 95% CI = [0.04, 0.09]; Fig. 3A), indicating that the congruency of the orientation of the object with that of the scene influenced performance on the orthogonal perceptual discrimination task. Supplementary analyses showed that congruency influenced both sensitivity and bias (**Fig. S4A**).

In this experiment, the object was congruent with the scene on a majority of trials. Real-world regularities (the coherence of an object's rotation with the surrounding scene), then, matched the short-term regularities observed during the experiment. In the next experiment, we investigated whether more frequent violations of real-world regularities would disrupt this effect of object congruence on task performance.

3.2. *Experiment 2: 50% probability*

In Experiment 2, the object reappeared from occlusion in a congruent or incongruent view with equal probability (i.e., 50% probability instead of 75%). Besides this probability manipulation, stimuli and experimental paradigm were the same as in Experiment 1. Like in the previous experiment, the included participants were reliably above chance in performing the task (mean accuracy and SEM: 0.69 ± 0.01).

Confirming the results of the previous experiment, in our main analysis we found accuracy to significantly differ between the Congruent and Incongruent conditions. Participants were more accurate on Congruent than Incongruent trials (means: 0.70 vs. 0.68, $t_{49} = 2.56$, $p = 0.01$, d = 0.45, 95% CI = [0.01, 0.05]; Fig. 3B). The effect of object congruency on task performance was thus consistent with that of Experiment 1 (results were also consistent for sensitivity and bias measures, see **Fig. S4B**). This suggests that even when the long-term expectation of scene and object rotating coherently was not informative of the stimuli shown during the experiment, it still affected participants' performance on the task. In Experiment 3, we asked whether presenting incongruent object orientations on a *majority* of trials could overrule, and possibly even reverse, the effect of object congruence on task performance.

3.3. *Experiment 3: 25% probability*

In this experiment, the object reappeared in the congruent orientation only on 25% of trials. Aside from this, stimuli and paradigm were the same as in the previous two experiments. Here, participants again performed well above chance level (mean accuracy and SEM: 0.69 ± 0.01).

In our central comparison of accuracy between the Congruent and Incongruent conditions, we again found a significant difference. Accuracy was higher in Congruent than Incongruent trials (means: 0.71 vs. 0.68, $t_{51} = 3.50$, $p < 0.001$, d = 0.47, 95% CI = [0.01, 0.05]; Fig. 3C), consistent with the previous experiments (results were also consistent for sensitivity and bias measures, see **Fig. S4C**). Interestingly, then, the influence of object congruence on task performance did not reverse when the short-term experimental regularities ran counter to it. This result provides strong evidence that the real-world constraint of coherent scene-object rotation cannot be easily overruled by inconsistent evidence within the short term of an experiment.

3.4. *Congruency-probability interaction*

The results of Experiments 2 and 3 revealed that participants predicted the orientation of the object to be coherent with the scene rotation, even when this was counter-productive for the task at hand. It is possible, however, that these expectations - that are derived from real-world regularities - still interact with the short-term contingencies observed during the experiment. In particular, when within-experiment and real-world regularities match, participants' expectations might be stronger than when they do not match.

To determine whether this was the case in our studies, we ran an across-experiment comparison (mixed ANOVA with Congruency as within-subject, Probability/experiment as between-subject factor) to test for the possible interaction of scene-object congruency and the probability of observing this congruency during the experiment. First, a main effect of Congruency confirmed that task performance was better on Congruent trials than Incongruent trials across experiments ($F_{1,\,149} = 49.31$, $p < 0.001$, $\eta_p^2 = 0.25$). Second, no main effect of Probability (experiment) on participant's accuracy was found, which shows that the staircasing procedure successfully yielded similar performance across experiments ($F_{2,\,149} = 0.57$, $p = 0.566$, $\eta_p^2 = 0.01$). Most interestingly, a significant interaction of Congruency and Probability on task performance was found ($F_{2,\,149} = 4.86$, $p = 0.009$, $\eta_p^2 = 0.06$). Subsequent two-sample (Welch) *t*-tests between the Congruent-Incongruent accuracy differences of different experiments revealed a significant difference between Experiment 1 (P(Congruent) = 75%) and both Experiments 2 and 3 (P(Congruent) = 50% and 25%; $t_{98} = 2.68$, $t_{92.7} = 2.52$, $p_{\mathrm{bonf}} =$ 0.02, 0.04, d = 0.54, 0.50 respectively). The accuracy differences in experiments 2 and 3, on the other hand, did not significantly differ ($t_{97.3} = -0.35$, $p_{\mathrm{bonf}} = 1.0$, d = 0.07)

The results of this analysis indicate that the impact of congruency on task performance was reduced when real-world contingencies did not reliably predict object appearance during the experiment. This suggests that observers use both long-term structural regularities and short-term experimental contingencies to predict object orientation.

3.5. *Role of alignment with the scene*

The spatial structure of scenes can be represented in two distinct ways: either in terms of image-like views from the observer's perspective (view-based representation) or in terms of object positions and orientations relative to the scene (structural representation). Cardinal axes of a scene (the axes parallel or perpendicular to walls) provide a reliable reference frame, that remains invariant to the observer's viewpoint. If the congruency effect that we observed is derived from a structural, scene-centric representation, then we would expect the congruency effect to be particularly strong when objects are aligned to these axes. If instead the congruency effect is derived entirely from a view-based, observer-centric representation, then the magnitude of the congruency effect should be unrelated to the alignment of the object with the main axes of the scene.

As the object orientations in our study (**Fig. S2**) could be divided into those that were aligned with one of the scene's main axes (0°, 180°) and those that were not (60°, 120°, 240°, 300°), we conducted an exploratory analysis to clarify the influence of object-scene alignment on our results. On the one hand, we tested whether the congruency effect is significantly stronger when object and scene are aligned, which would indicate a role of structural cues. On the other hand, we tested whether the congruency effect is still present on misaligned trials, which would show that structural cues, even if they contribute to scene-driven expectations, are not necessary for observers to form such expectations.

First, to examine the effects of Congruency, Alignment, and Probability (experiment) together, as well as possible interactions between them, we conducted a mixed ANOVA with Congruency and Alignment as within-subject factors, and Probability as a between-subject factor. The full results of the analysis are reported in Table 1. Beyond the effects of Congruency and Probability, already reported in the main text, this analysis revealed a significant main effect of Alignment (mean accuracies: 0.72 and 0.67 for Aligned and Misaligned objects respectively; $F_{1,149} = 57.98$, $p < 0.001$, $\eta_p^2 = 0.27$), consistent with the established finding that participants are more sensitive to orientation differences around cardinal axes (Appelle, 1972; Shiffrar & Shepard, 1991). More importantly, the interaction between Congruency and Alignment was also significant ($F_{1,151} = 5.13$, $p = 0.025$, $\eta_p^2 = 0.03$), showing that the accuracy difference between Congruent and Incongruent trials was

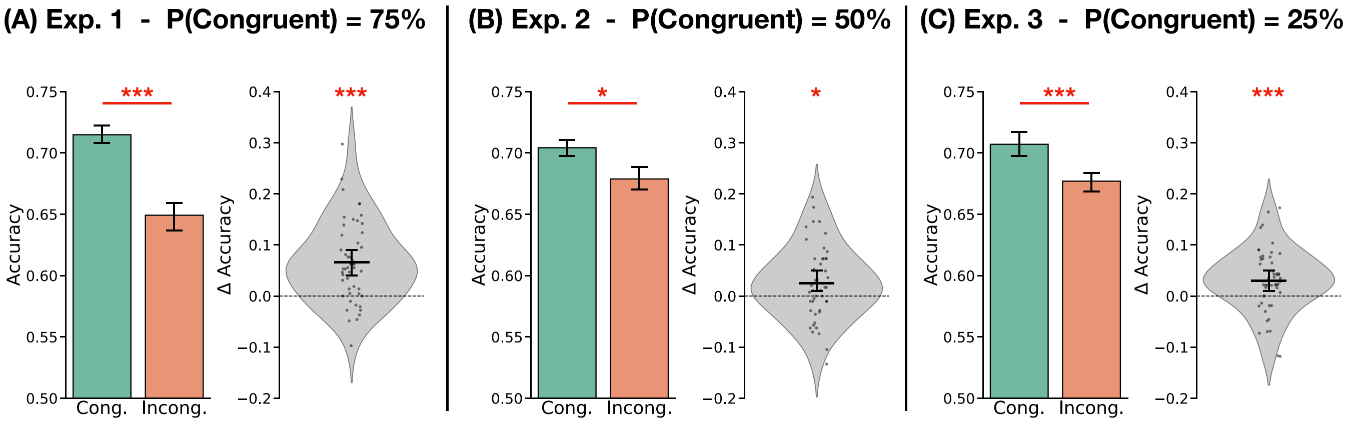

Fig. 3. Results of Experiments 1–3. Bar plots show mean accuracy (and SEM) for Congruent and Incongruent trials. Violin plots show distribution of the differences between conditions (Congruent – Incongruent) for each participant, with mean and 95% confidence interval. * $p < 0.05$, *** $p < 0.001$.

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

Effect	df	F	p	η_p^2
Congruency	1, 149	**51.01**	**< 0.001*** **	**0.25**
Alignment	1, 149	**57.98**	**< 0.001*** **	**0.27**
Probability	2, 149	0.53	0.588	0.01
Congruency x Alignment	1, 149	**5.13**	**0.025** *	**0.03**
Congruency x Probability	2, 149	**4.35**	**0.015** *	**0.05**
Alignment x Probability	2, 149	**3.70**	**0.027** *	**0.05**
Congr. x Align. x Prob.	2, 149	0.08	0.920	0.00

larger when objects were aligned than when objects were misaligned with the main axes of the scene (mean accuracy differences: 0.06 and 0.03 respectively). Structural cues, then, might have contributed to participants' expectations of the upcoming object orientation.

Next, we asked whether the congruency effect was still present when object and scene were misaligned, to determine whether structural cues are *necessary* for object expectations to emerge. To this end, we separately tested the difference between Congruent and Incongruent trials in the Aligned and Misaligned conditions. As the ANOVA did not reveal a three-way interaction between Congruency, Alignment and Probability (i.e., the interaction between congruency and alignment did not differ between experiments; Table 1), we grouped the three experiments together to maximize statistical power. Here, we found that the effect of Congruency was significant for both Aligned ($t_{151} = 5.33$, $p < 0.001$, $d = 0.59$, 95% CI = [0.04, 0.08]) and Misaligned ($t_{151} = 4.16$, $p < 0.001$, $d = 0.41$, 95% CI = [0.02, 0.04]) trials, as shown in Fig. 4. The congruency effect, then, was still present when object and scene were misaligned.

Altogether, these results indicate that structural cues were not necessary for participants' scene-driven expectations of object orientation to arise. As our stimuli did not include other apparent structural cues,[1] it is likely that on misaligned trials object expectations were elicited by view-based cues. On the other hand, the congruency effect was significantly stronger on aligned trials, suggesting that structural cues might have also contributed to participants' expectations. It might not be possible, however, to rule out that other aspects of our experimental design might have contributed to this interaction: (1) the lower general accuracy on misaligned trials might have led to a floor effect, attenuating the difference between congruent and incongruent trials; (2) in the aligned trials, incongruent objects were also misaligned with the scene, possibly accentuating the effect of congruency.

While it might not be possible, then, to conclusively show using the present analysis that structural scene cues played a role in our paradigm, previous findings and theoretical motivations suggest that the representations involved were not purely view-based, as we outline in the **Discussion**.

4. Discussion

In the real world, objects and their context are strongly interdependent, allowing observers to predict how the orientation of an object will change based on changes in the orientation of the surrounding scene. In this study, we manipulated whether objects respected this constraint, and measured how this affected participants' performance on an orthogonal perceptual discrimination task. Across three experiments, we found that participants' accuracy was influenced by scene-object congruency, showing that they formed an expectation of the object's updated orientation based on the surrounding scene.

The scene-object orientation congruency effect revealed here could not be driven by physical stimulus differences between conditions, as the (in)congruency of a given object orientation was determined exclusively by the trial context: the object's initial orientation, and the rotation of the scene background while the object was occluded. Moreover, the effect appeared to occur automatically: (1) the task did not require predicting object orientation, or even taking scene information into account; (2) as revealed by our post-experiment survey, the congruency effect did not correlate with participants' self-reported prediction of the object's orientation, nor with self-reported attention to the scene context (see Supplementary Materials **S.5**); (3) most importantly, the effect was still present even when the real-world constraint was not predictive (Experiment 2) or even was counter-predictive (Experiment 3) during the experiment. In Experiments 2 and 3, predicting the congruent object orientation would not help performing the task, so it is particularly unlikely that participants predicted object orientation as part of a deliberate strategy.

While short-term experiment regularities did not erase long-term structural expectations we still found that the accuracy difference was significantly greater in Experiment 1 than both Experiments 2 and 3. As such, participants may have been able to partly suppress real-world expectations when they ran counter to the current task setting (Dogge, Custers, Gayet, Hoijtink, & Aarts, 2019). The fact that the scene-object congruency effect did not reverse in Experiment 3 (where real-world contingencies were counter-predictive) further indicates that scene-based predictions of object appearance were not solely driven by

[1] Another possible cue is the orientation of the main object relative to background objects: however, these were aligned with the scene's walls, meaning that this cue could also not have been used on misaligned trials.

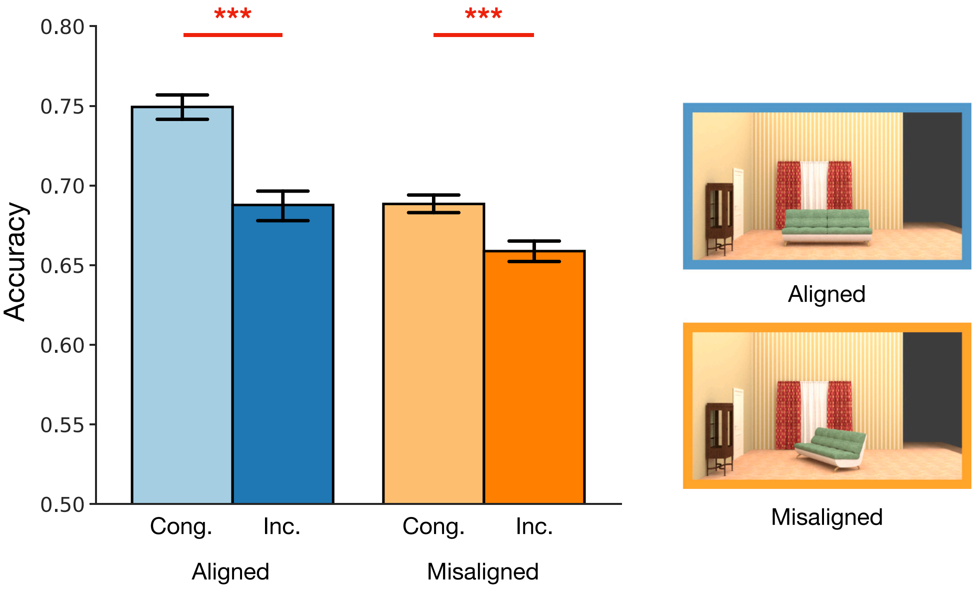

Fig. 4. Mean accuracy (and SEM) for Congruent and Incongruent trials, plotted separately for Aligned and Misaligned objects. *** $p < 0.001$.

statistical association between arbitrary object views, unlike in prior studies investigating the effect of probabilistic expectations on perception (e.g., Meyer & Olson, 2011; Richter, Ekman, & de Lange, 2018). Instead, scene-based predictions of object appearance were constrained by real-world regularities, acquired through life-long learning. Furthermore, these expectations were flexible: both the amount of overall scene rotation and the orientation of the object relative to the scene were varied across trials, meaning that participants were able to predict objects from novel viewpoints given any initial orientation, and to adjust this prediction to the amount of rotation in the scene.

The present findings show that scene representations can support object predictions across spatial transformations. What could be the format of these scene representations? One possibility is that the scene is represented in a viewpoint-invariant way, in terms of its parts and their spatial relations (Biederman, 1987; Erdogan & Jacobs, 2017; Hummel, 2000). In that case, the incongruency of an object view would correspond to the detection of a change in the object's orientation relative to the scene. Alternatively, the scene may be represented as a collection of image-like views, linked by operations such as mental rotation (Tarr & Pinker, 1989), combination (Ullman, 1998), associative learning (Gillner & Mallot, 1998; Glennerster, 2023; Gootjes-Dreesbach, Pickup, Fitzgibbon, & Glennerster, 2017) or normalization (Willems & Wagemans, 2001). Expectation violations, in that case, would mismatch the egocentric object view predicted by the participant.

To tentatively distinguish between these two accounts, in an exploratory analysis we compared the magnitude of the congruency effect between trials in which the object was aligned with one of the cardinal axes of the scene, and trials in which it was not. This comparison is commonly used to adjudicate between structure- and view-based scene representations (e.g., Marchette & Shelton, 2010; Mou & McNamara, 2002), since salient axes of the environment provide a stable reference frame that can be used across viewpoints. We found that the congruency effect was still present when the object was not aligned with the scene, which is consistent with the view that scene-driven orientation expectations can be elicited by view-based cues alone (for similar conclusions using a traditional mental rotation paradigm, see Stewart et al., 2022). On the other hand, our finding of a slightly larger congruency effect for aligned than for misaligned objects might indicate an additional contribution of structural scene cues. While it might not be possible to determine this from this analysis alone, as described in the **Results**, prior findings and theoretical motivations suggest that view-based and structural information were likely to be both involved. Several studies have shown that humans extract these two types of information in parallel, both in object perception and spatial cognition (Burgess, Spiers, & Paleologou, 2004; Foster & Gilson, 2002; Heywood-Everett, Baker, & Hartley, 2022). From a computational standpoint, several hybrid models have been proposed, in which spatial relations are represented in an approximate way, remaining bound to viewer-centered image coordinates (Bear et al., 2020; Edelman & Intrator, 2001). Such hybrid models are a compromise between the compositional flexibility afforded by structural representations (Hafri, Green, & Firestone, 2023) and the necessity to estimate them from egocentric views, as well as the potential advantage of using retinotopic coordinates as a common reference frame for different visual computations (Groen, Dekker, Knapen, & Silson, 2022). An intriguing question for future research is the extent to which these two objectives are balanced in the scene representations underlying dynamic object tracking and prediction.

The present results complement previous findings showing that the way objects are perceived and represented depends on static scene context. While previous work primarily focused on how scene context facilitates recognition of an object's identity (Bar, 2004; Brandman & Peelen, 2017; Oliva & Torralba, 2007; Võ et al., 2019), a property that remains invariant to viewpoint changes, we here investigate object *orientation*, a property that is inherently dependent on viewpoint changes.

Other studies have shown that depth cues in a scene, indicating the distance of an object from the observer, can influence the object's perceived size, as in the classic Ponzo illusion (Leibowitz, Brislin, Perlmutrer, & Hennessy, 1969; Yildiz, Sperandio, Kettle, & Chouinard, 2021). Interestingly, beyond shaping the representation of perceived objects, scene depth cues can also 'rescale' internally generated object representations, such as preparatory templates in visual search (Gayet & Peelen, 2022). Unlike these previous studies, however, here we demonstrate that internal object representations can be influenced by scenes dynamically, being updated as the scene changes. Moreover, participants in our study were not engaged in an explicit visual search task, showing that object representations automatically transform in accordance with the scene context.

Scene context can also provide cues to 3D object orientation: the difficulty of recognizing objects in unfamiliar orientations can be alleviated if scene context provides a spatial reference frame (Christou, Tjan, & Bülthoff, 2003; Humphrey & Jolicoeur, 1993). In the present study, we show that this contextual information can also drive internal predictions of object appearance from new viewpoints, consistent with the possibility that similar mechanisms might underlie context-driven transformations and the mental rotation of isolated objects (Graf, 2006). Future research could further investigate the relation between these cognitive processes, elucidating how internal representations and external context interface in real-world perception.

5. Conclusion

In conclusion, we have shown that participants create expectations of object appearance from novel viewpoints automatically, driven by scene context. These expectations affect participants' accuracy in an orthogonal perceptual discrimination task. Moreover, the expectations that are based on real-world contingencies are not easily overruled by frequent violations, and persist even when detrimental to task performance, showcasing their automaticity. Together, our results demonstrate that scene context facilitates the mental transformation of objects, supporting efficient perception in structured real-world environments.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105521.

Funding

This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 725970).

CRediT authorship contribution statement

Giacomo Aldegheri: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. **Surya Gayet:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Marius V. Peelen:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Data availability

All data and stimuli are publicly available at https://osf.io/wnefh/. Code for running the online experiments and analyzing the data is publicly available at https://github.com/GAldegheri/scenecontext-transforms. The design and analysis plans for the experiments were not preregistered.

References

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin, 78*(4), 266.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617–629.

Bear, D., Fan, C., Mrowca, D., Li, Y., Alter, S., Nayebi, A., … Others. (2020). Learning physical graph representations from visual scenes. *Advances in Neural Information Processing Systems, 33*, 6027–6039.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115.

Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *Journal of Neuroscience, 37*(32), 7700–7710.

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

Burgess, N., Spiers, H. J., & Paleologou, E. (2004). Orientational manoeuvres in the dark: Dissociating allocentric and egocentric influences on spatial memory. *Cognition, 94* (2), 149–166.

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

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences, 22*(9), 764–779.

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

Dogge, M., Custers, R., Gayet, S., Hoijtink, H., & Aarts, H. (2019). Perception of action-outcomes is shaped by life-long and contextual expectations. *Scientific Reports, 9*(1). https://doi.org/10.1038/s41598-019-41090-8. Article 1.

Edelman, S., & Intrator, N. (2001). A productive, systematic framework for the representation of visual structure. *Advances in Neural Information Processing Systems*, 10–16.

Erdogan, G., & Jacobs, R. A. (2017). Visual shape perception as Bayesian inference of 3D object-centered shape representations. *Psychological Review, 124*(6), 740.

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society B: Biological Sciences, 269*(1503), 1939–1947. https://doi.org/10.1098/rspb.2002.2119

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

Gillner, S., & Mallot, H. A. (1998). Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of Cognitive Neuroscience, 10*(4), 445–463.

Glennerster, A. (2023). Understanding 3D vision as a policy network. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 378*(1869), 20210448. https://doi.org/10.1098/rstb.2021.0448

Gootjes-Dreesbach, L., Pickup, L. C., Fitzgibbon, A. W., & Glennerster, A. (2017). Comparison of view-based and reconstruction-based models of human navigational strategy. *Journal of Vision, 17*(9), 11.

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

Groen, I. I. A., Dekker, T. M., Knapen, T., & Silson, E. H. (2022). Visuospatial coding as ubiquitous scaffolding for human cognition. *Trends in Cognitive Sciences, 26*(1), 81–96. https://doi.org/10.1016/j.tics.2021.10.011

Hafri, A., Green, E. J., & Firestone, C. (2023). Compositionality in visual perception. *PsyArXiv*. https://psyarxiv.com/trg7q.

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., … Smith, N. J. (2020). Array programming with NumPy. *Nature, 585*(7825), 357–362.

Heywood-Everett, E., Baker, D. H., & Hartley, T. (2022). Testing the precision of spatial memory representations using a change-detection task: Effects of viewpoint change. *Journal of Cognitive Psychology, 34*(1), 127–141.

Higgins, I., Racanière, S., & Rezende, D. (2022). Symmetry-based representations for artificial and biological general intelligence. *Frontiers in Computational Neuroscience, 16*. https://www.frontiersin.org/articles/10.3389/fncom.2022.836498.

Hummel, J. E. (2000). *Where view-based theories break down: The role of structure in shape perception and object recognition* (pp. 157–185). Cognitive Dynamics: Conceptual Change in Humans and Machines.

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.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*(4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265–270.

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

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. (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

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

Marchette, S. A., & Shelton, A. L. (2010). Object properties and frame of reference in spatial memory representations. *Spatial Cognition and Computation, 10*(1), 1–27. https://doi.org/10.1080/13875860903509406

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

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences, 108*(48), 19401–19406. https://doi.org/10.1073/pnas.1112895108

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 162.

Munton, J. (2022). How to see invisible objects. *Noûs, 56*(2), 343–365.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11*(12), 520–527.

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

Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *Journal of Neuroscience, 38*(34), 7452–7461.

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

Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review, 91*(4), 417.

Shepard, R. N. (2001). Perceptual-cognitive universals as reflections of the world. *Behavioral and Brain Sciences, 24*(4), 581–601.

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

Shiffrar, M. M., & Shepard, R. N. (1991). Comparison of cube rotations around axes inclined relative to the environment or to the cube. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 44–54. https://doi.org/10.1037/0096-1523.17.1.44

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

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21*(2), 233–282.

Ullman, S. (1998). Three-dimensional object recognition based on the combination of views. *Cognition, 67*(1–2), 21–44.

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., … Bright, J. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods, 17*(3), 261–272.

Võ, M. L.-H., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology, 29*, 205–210.

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.

Willems, B., & Wagemans, J. (2001). Matching multicomponent objects from different viewpoints: Mental rotation or normalization? *Journal of Experimental Psychology: Human Perception and Performance, 27*(5), 1090.

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.

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.