Spatial correspondence in relative space regulates serial dependence

Jaeseob Lim[1] **and Sang-Hun Lee**[1, *]

[1]Department of Brain and Cognitive Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea (08826)
[*]Corresponding author (email: visionsl@snu.ac.kr)

Abstract

Our perception is often attracted to what we have seen before, a phenomenon called 'serial dependence (SD).' From a normative standpoint, SD can help us maintain a stable perception of the world by considering the statistical regularity in the environment. To serve this presumed normative function, SD must occur more strongly between consecutive elements sharing the same identity than between those with different identities when multiple elements undergo dynamic shifts in position across successive views. However, what factors contribute to such selective SD between identity-matching elements in dynamic situations remains unknown. Here, we hypothesized that SD preferentially occurs between consecutive elements matched in relative rather than absolute space because spatial correspondence in relative coordinates can warrant identity matching invariantly to changes in absolute coordinates. To test this hypothesis, we developed a task where two targets change positions in unison between successive views. The results supported our hypothesis: SD was stronger when consecutive targets matched in relative coordinates than in absolute coordinates. Moreover, SD transmission in relative coordinates was spatially more specific than previously observed in absolute coordinates. The characteristics of SD we found here promote object-based perceptual stabilization over time in dynamic situations.

Introduction

Objects in the natural setting rarely alter their features. Even when alterations do occur, they are not random but display statistical regularities over time [1,2]. For example, sequential items are likely to have similar orientations [3]. Given this temporal steadiness of the natural environment, an ideal perceptual system would exploit this stability to accurately estimate the features of objects. One intriguing phenomenon supporting this perspective [2,4] is the bias in estimation, called "serial dependence (SD)," where our perceptual estimates tend to gravitate to the previous estimates [5,6]. SD is known to become stronger as the spatial distance between successive targets decreases [5,7,8]. This spatiotemporal specificity of SD, referred to as the "continuity field [5]," gives it a place in the discussion of the perceptual system's utilization of environmental stability because the continuity field appears to reflect the statistical regularities in the natural environment [2,4].

In the natural setting, our views consist of many objects or groups of items. In such a situation, to benefit from the temporal stability of the environment effectively, the perceptual system must solve the "identity-matching problem," determining which elements of a single object or components within a group correspond to which parts or components between successive views [9]. If such elements or components are mismatched, SD can be detrimental to the system, as it invites additional variability to perceptual estimates by attracting them to other irrelevant (i.e., non-identical) elements or components in the past view. Previous studies suggest that SD can be routed between sequential items such that identity matching occurs based on non-target features of objects or contextual information [8,10–13]. In this respect, the spatial correspondence of sequential elements provides important contextual information that helps solve the identity-matching problem, indicating the likelihood that any successive elements share the same identity. For example, sequential items are more likely to share the same identity when they appear in the same location than in different locations (Fig. 1).

For a comprehensive understanding of how the perceptual system utilizes spatial correspondence in solving the matching problem for SD, it is critical to distinguish between the spatial proximity defined in absolute and that defined in relative coordinates. In absolute coordinates, the locations of elements are defined in reference to the fixed points in the world or retinotopic space (e.g., the meridians). In relative coordinates, the locations of elements are defined in reference to the other elements or points within an object (e.g., the eyes within a face) [14–18]. It is essential to note that successive elements' spatial locations and proximity depend on the position of 'other' elements in relative coordinates but not in absolute coordinates. The distinction between proximity in absolute and relative coordinates becomes critical when objects or groups of elements change their positions over successive views (Fig. 1b,c), as is often the case in the natural setting. When dealing with the identity-matching problem in such non-static situations, the

perceptual system should rely on spatial proximity in relative coordinates to effectively exploit the temporal stability through SD.

The rationale presented above suggests that SD should occur preferentially between successive elements matched in relative coordinates as opposed to in absolute coordinates when objects or groups of elements change their locations over successive views. Besides this normative rationale, prior empirical studies on cognitive acts other than SD suggest that human observers use the spatial proximity defined in relative coordinates in similar dynamic situations. In a letter naming task, the priming effect in naming speed was determined by whether a target matched a primer in relative coordinates instead of absolute (retinotopic) coordinates [19]. In a visual search task, the pre-cueing effects, both in speed and accuracy, were also governed by the spatial proximity in relative coordinates [20]. Where motion perception tasks require the temporal integration of multiple successive elements, the motion integration path conformed to the spatial proximity defined in relative coordinates [21,22]. Note that the tasks used in these studies, despite their differences in stimuli and responses, commonly entail the cognitive operation of identity matching—determining which elements match which elements over successive views—in non-static situations. Given that SD also entails the identity-matching problem, as noted earlier, human reliance on the proximity defined in relative coordinates in the naming, search, and motion-perception tasks implicates that SD is likely to transpire over the successive elements matched in the relative coordinates when multiple elements undergo a shift as a group or as parts within an object.

Despite the normative basis (Fig. 1) and empirical plausibility of routing SD based on the proximity in relative coordinates, previous studies on SD have investigated the spatial tuning property of SD only in absolute coordinates [5,7,23]. The lack of exploration into the spatial specificity of SD in relative coordinates is a consequence of earlier SD research relying on task paradigms that presented a single element in each view [6,9] or pre-cued only a single target even when presenting multiple elements [5,24,25]. Multiple 'task-relevant' elements are required to define the locations and proximities of successive elements in relative coordinates. For that matter, a recent study [8] reported the spatial specificity of SD in a setup with two task-relevant targets displayed at each view; however, such specificity cannot be considered to signify the spatial specificity of SD in relative coordinates since the two target locations remain fixed over successive views (, like those shown in Fig. 1a). The spatial proximities in relative and absolute coordinates are distinguished only when multiple elements shift locations over successive views (Fig. 1b,c).

Here, we developed a task paradigm where two oriented line elements shift their locations in unison over successive trials while maintaining their distance, and observers estimated the orientation of the retro-cued element on each trial (Fig. 2a,b; Fig. 3a,b). Following the previous work that demonstrated the effect of the proximity in relative coordinates on motion perception [21,22,26], we promoted the grouping

of the two elements, so that they are treated as two components constituting a single object, by leveraging the 'proximity between elements' and 'common fate' rules of Gestalt [27–29]. With this task paradigm, we show that the orientation estimate of the current target is more strongly attracted to the orientation of the previous target near it in relative coordinates than in absolute coordinates. This suggests that human observers rely on spatial proximity in relative coordinates in routing SD between multiple elements over successive views to exploit environmental stability effectively.

Results

Experiment 1

Fifteen observers participated in Experiment 1 at home via the Internet. To control the spatial scale of the display on this web-based platform, we instructed observers to keep a distance of 60 cm from their eyes to the monitor and report the size of the display using their credit card as a reference (see Methods for details). On each trial, while fixating on a cross, observers briefly viewed two gratings placed one above the other and separated by a distance of 6 cm on the right side of the view (12 cm away from the fixation). After the two gratings disappeared and were masked, and after a delay, we showed an arrow at the fixation spot to indicate the location of a target grating retrospectively. Observers then reproduced the orientation of the target grating by rotating a pair of dots on the circle around the fixation (Fig. 2a).

To be reminded, our goal is to determine whether the proximity defined in absolute coordinates or the one defined in relative coordinates is what determines how SD is routed over multiple elements over successive views. As one effective way to achieve this goal, we vertically shifted the two gratings up and down together—while keeping their distance (6 cm)—by 4 cm between two fixed positions (Fig. 2b). With this manipulation, we could compare the strength of SD in the two proximity conditions. In the 'same-relative-location' condition (the inset in the blue box of Fig. 2c), the previous and current targets are matched in relative coordinates but substantially (by 4 cm) away from each other in absolute coordinates. In the 'different-relative-location' condition (the inset in the red box of Fig. 2c), the previous and current targets are not matched in relative coordinates but near (apart by 2 cm) to each other in absolute coordinates.

It should be noted that we intentionally did not fully randomize the placement of the retro cue in successive views. Under complete randomization, there would be four possible cases of retro-cue transition for both proximity conditions. We chose to exclude two cases from the different-relative-location condition where the previous and current targets were positioned on opposite edges and were too far apart (10 cm). Also, we modified the probabilities of the remaining cases to equate the number of trials between the two proximity conditions (see Methods for details). As a result, the distance in absolute coordinates between successive targets was consistently shorter in the different-relative-location condition

than in the same-relative-location condition. This fact allows us to safely ascribe any higher degree of SD in the same-relative-location condition to the superlativeness of the proximity in relative coordinates to that in absolute coordinates as a factor routing SD. This is because SD must be stronger in the different-relative-location condition than in the same-relative-location condition if the proximity in absolute coordinates determines how SD is routed over multiple elements.

We quantified the strength of SD and compared it between the two proximity conditions in the following steps. First, for each condition, we took the averages of the errors in orientation estimates across observers and derived an SD curve by plotting those averaged errors as a function of the orientation difference between the previous and current targets (thin lines with light shades in Fig. 2c). Second, after confirming that our data showed a horizontal S shape similar to the shape of SD reported in previous studies, we proceeded to fit the Gaussian derivative function to the observed SD curves (thick curves in Fig. 2c). Finally, we quantified the strength of SD with the amplitude of the best-fit Gaussian derivative function (vertical arrows in Fig. 2c).

We found that the amplitude of the SD curve was greater in the same-relative-location condition (4.416°; 95% CI, [3.473°, 5.489°]) than in the different-relative-location condition (2.627°; 95% CI, [1.613°, 3.869°]) (permutation test, $p = 0.022$). These findings indicate that SD occurred preferentially between the targets matched in relative position even though they were far apart in absolute position.

Experiment 2

Experiment 1 suggests that the spatial proximity defined in relative coordinates routes SD between multiple elements over successive views. However, its results can be interpreted in different ways due to two aspects of its experimental design. First, we were unable to ascertain whether observers maintained their gaze on the fixation point as instructed because the gaze was not monitored. One problematic scenario is that observers could have shifted their gaze near the gratings, thus moving it up and down in synchrony with the gratings, to benefit from foveal vision. If so, the retinotopic distance between successive targets would consistently be shorter in the same-relative-location condition than in the different-relative-location condition. Therefore, the observed heightened level of SD in the same-relative-location condition could have resulted from the closeness of consecutive targets in retinotopic coordinates [7]. Second, the two gratings alternatingly appeared in the two fixed positions, while the locations of targets were not fully randomized in Experiment 1, as pointed out earlier. Observers could have utilized this spatiotemporal contingency to bias their attention to the grating with a higher chance of being retro-cued as a target.

We conducted Experiment 2 with a different spatial layout (Fig. 3a,b) to confirm the findings of Experiment 1 while addressing the two alternative interpretations described above. Unlike in Experiment 1, twenty-four observers visited our laboratory to participate in Experiment 2 under a highly controlled environment. On each trial, two gratings that were horizontally arranged below the fixation cross randomly appeared on either the left side (marked as 'X' and 'Y' in Fig. 3a) or the right side (marked as 'Y' and 'Z' in Fig. 3a) of the three retinotopically fixed locations. To prevent any spatiotemporal contingency between successive targets, we completely randomized the placement of targets within a given pair of gratings across trials. To ensure that observers' gaze was steady, we monitored their eye movements (see Methods for details). Trials where the gaze deviated by over 1.5 degrees in visual angle (dva) during the fixation period were removed from further analysis.

The spatial arrangement and complete randomization of targets in Experiment 2 allowed us to manipulate the spatial proximity of successive targets separately for the absolute and relative coordinates. The successive targets were apart by 0, 7, or 14 dva in the absolute coordinates and by 0 and 7 dva in the relative coordinates (Fig. S1). To match the levels of proximity between the absolute and relative coordinates, we excluded the trials where the successive targets were 14 dva apart (marked as dashed boxes in Fig. S1) from further analysis. As a result, the remaining trials could be sorted into two-by-two factorial conditions based on absolute and relative coordinates (Fig. 3b).

To determine if SD routes between successive targets based on proximity in relative coordinates—as observed in Experiment 1, we first contrasted the proximity conditions in terms of the relative coordinates by merging the trials differing in the proximity in absolute coordinates. We then estimated the amplitudes of SD in the same method used in Experiment 1, separately for the same-relative-location (0 dva apart; cells outlined by a blue box in Fig. 3b) and different-relative-location (7 dva apart; cells outlined by a red box in Fig. 3b) conditions. Results were consistent with those from Experiment 1: the amplitude of the SD curve was greater in the same-relative-location condition (4.142°; 95% CI, [3.479°, 4.887°]) than in the different-relative-location condition (2.659°; 95% CI, [1.925°, 3.420°]) (permutation test, $p = 0.004$).

Note that the randomization of the stimulus and target locations resulted in an unequal number of trials between the same-relative-location and different-relative-location conditions in the marginal comparison above (Fig. S3). This imbalance invites a potential confounding variable because the same-relative-location and different-relative-location conditions differ not only in the proximity in relative coordinates but also in the trial proportions of the same-absolute-location and different-absolute-location conditions. In other words, it is possible that the stronger SD in the same-relative-location condition (Fig. 3c) could have resulted from the fact that the proportion of the same-absolute-location trials was greater in the same-relative-location condition than in the different-relative-location condition. To check this

possibility, we carried out conditional comparisons where the same-relative-location and different-relative-location conditions were compared in SD while keeping constant the proximity in absolute coordinates (Fig. 3d). We found that the amplitude of SD consistently tended to be greater in the same-relative-location condition than in the different-relative-location condition regardless of whether the successive targets were matched (one-tailed permutation test, p=0.087; circles with a dashed line in Fig. 3d) or not (one-tailed permutation test, p=0.0002; squares with a solid line in Fig. 3d) in absolute coordinates.

The results of the conditional comparisons (as seen in Fig. 3d) rule out the possibility that the pronounced SD in the same-relative-location condition observed in the marginal comparison (as seen in Fig. 3c) is due to the difference in the trial proportions of the same-absolute-location and different-absolute-location conditions. This possibility is further disproven by the marginal comparison in SD between the same-absolute-location (the top row in Fig. 3b) and different-absolute-location (the bottom row in Fig. 3b) condition. To our surprise, the amplitude of SD tended to be greater in the different-absolute-location condition (3.853; 95% CI, [3.258, 4.565]) than in the same-absolute-location condition (2.939; 95% CI, [2.268, 3.678]) (permutation test, p=0.07).

Spatial specificity of the SD associated with the previous non-target orientation

Our retro-cueing procedure requires observers to retain the orientation information for both the target and non-target stimuli during the delay period. Though not overtly reported, the retained orientation of non-target stimuli in the previous trial may still impact how the current trial's target orientation is perceived or remembered. This suggests the possibility that SD might also occur between the non-target orientation in the previous trial and the reported target orientation in the current trial, which will be referred to as 'non-target SD.' This possibility prompts a question of whether the proximity in relative coordinates would still regulate the strength of the non-target SD.

To answer this question, we quantified the non-target SD for the dataset acquired in Experiment 2 and compared its strength between the same-relative-location and different-relative-location conditions. In the same-relative-location (blue curve in Fig. 4), the non-target SD was substantial (1.89; 95% CI, [1.20, 2.87]) and significantly different from zero (permutation test, p = 0.005). By contrast, the non-target SD was absent in the different-relative-location condition (-0.19; 95% CI, [-1.72 ~ 1.23]; permutation test, p = 0.572). The difference between the two conditions was statistically significant (permutation test, p = 0.002). A similar trend of results was observed for the dataset in Experiment 1, though not statistically significant (Fig. S2): the non-target SD was stronger in the same-relative-location condition (1.714; 95% CI, [0.657, 3.047]) than in the different-relative-location condition (0.580; 95% CI, [-6.192, 2.525]) (permutation test,

p = 0.267). These results suggest that SD preferentially occurs between the successive elements matched in relative coordinates as long as the previous element's feature is retained in memory, irrespective of whether it was overtly reported or not.

Discussion

Guided by the view that SD reflects the cognitive act of leveraging environmental stability to estimate object features, we hypothesized that SD selectively occurs between successive elements matched in relative coordinates when multiple elements shift in position over time. Consistent with this hypothesis, we found that the spatial correspondence in relative coordinates regulated SD. Specifically, SD substantially decreased as the distance between the successive targets increased in relative coordinates. By contrast, such a distance-dependent decrease in SD was not found when the distance between successive targets was defined in absolute coordinates, following the way the spatial specificity of SD has been conventionally defined [5,7,23,30].

There are two different perspectives to consider when explaining how the spatial location of elements impacts SD depending on whether it relates to SD based on local features or element identities. The 'feature-based' view focuses on the statistical regularity governing the spatiotemporal relationship among local features in the environment. In this view, as consecutive elements get closer in space, they tend to have similar features, which leads to an increase in SD. Here, the spatial proximity between consecutive elements directly regulates SD, irrespective of their identity being the same or different. The feature-based view is well captured by the concept of the "continuity field" [5], which assumes that the perceptual system strives to achieve visual perception with spatiotemporal stability in terms of local features, where element identity does not play any role. By contrast, in the 'identity-based' view, the spatial proximity between consecutive elements indicates the likelihood that they share the same identity. Here, SD is expected to be pronounced for the consecutive elements close to each other because they are likely to belong to the same entity. In other words, spatial proximity can also function as a cue for solving the identity-matching problem in SD. Once again, the term "identity-matching problem" that we used here refers to the process of determining which element matches with which element among multiple elements. So, the identity-based view could not be explicitly addressed in the earlier work investigating the effects of spatial location on SD because only a target element was presented alone or pre-cued at each view [5,7,23,30]. The identity-matching problem does not exist in such single-target situations, and the feature-based view alone can sufficiently explain the spatial specificity of SD. By contrast, the SD should be selectively formed between identity-matched elements for object-based stabilization in a situation

where the "multiple potential targets" comprise each view like in our experiments (Fig. 2). Importantly, by separating two kinds of spatial location information, one critical for identity matching (location in relative coordinates) and the other irrelevant to identity-matching but more fundamental and carrying low-level spatial information (location in absolute coordinates), our study could demonstrate that the former type of spatial information regulates SD when multiple potential targets shift in position over consecutive views.

We only moderately varied the distance between consecutive targets in relative coordinates: 0 dva (same-relative-location) or 5.7 dva (different-relative-location) in experiment 1; 0 dva or 7 dva in experiment 2. Despite such small offsets, SD significantly decreased in magnitude when the consecutive targets were mismatched in relative coordinates (39% in exp 1 and 36% in exp 2). These results are hard to explain in terms of the spatial specificity of SD reported in previous studies, where SD did not decrease as the consecutive targets became apart by 10 dva in absolute coordinates [24] and began to decrease substantially after the consecutive targets became apart by farther than 10 dva in absolute coordinates [5,7]. The differences in spatial specificity of SD between the previous studies and ours should be ascribed not to other nuisance variables, such as stimulus contrast and frequency, but to whether the distance between targets is defined in absolute or relative coordinates. This is because when our data were sorted in terms of the distance between consecutive targets in absolute coordinates, we could not find any difference in SD strength between 0 dva and 7 dva. Thus, the spatial specificity of SD seems higher when defined in relative coordinates than in absolute coordinates.

This high spatial specificity of SD in relative coordinates is consistent with the view that relative spatial location functions as a cue for solving the identity-matching problem when multiple potential targets comprise each view and change their locations between consecutive views. In such multiple-target situations, the targets can be located quite closely in space, as in our study. For SD to selectively occur between the elements sharing the same identity, the identity-matching problem must be solved even for such a narrow range of distance between simultaneous targets. However, when only a single target exists at each view, as in previous studies, two consecutive targets are likely to share the same identity even if they are far apart [7]. Hence, the spatial specificity of SD is expected to be higher when multiple potential targets are present at each view than when a single target is at each view.

SD is known to occur also for complex features such as facial expression [10,12,13,31]. A straightforward account of this phenomenon is that SD occurs directly between successive high-level feature representations. For example, in the case of facial emotion expression, SD may occur because the current 'emotional state' is attracted to the previous 'emotional state.' This 'high-level' account separates the SD in high-level features from the SD in low-level local features comprising the high-level features,

such as spatial frequency [32], orientation [5,24,33], and color [34]. The high spatial specificity of SD in relative coordinates observed in our study makes it possible to offer an alternative—but not exclusive to the high-level account—bottom-up account of the SD in complex features. In the bottom-up account, an ensemble of local elements (e.g., the angles of the left and right eyebrows and mouth) in the current view assimilates to that in the previous view, while the state of a high-level feature (e.g., 'happy' facial expression) is inferred from the ensemble of local features at any given view. As a result, the assimilation of the current feature ensemble to the previous feature ensemble makes the current state of the high-level feature in the current view attracted to the previous state of the high-level feature inferred in the previous view. Importantly, the prerequisite of the bottom-up account is a highly selective identity-matched SD over the multiple local elements between consecutive views. For example, the angle of the left eyebrow in the current view must assimilate not to that of the right eyebrow but to that of the corresponding left eyebrow in the previous view. Otherwise, SD can be detrimental to the stable perception of facial expressions. Thus, the bottom-up account of the SD for high-level features is not plausible if the spatial specificity of SD is only broadly regulated in absolute coordinates, as known previously. Instead, the special specificity of SD seems to be sharply regulated in relative coordinates, as our work implies, given that the SD for high-level features can occur even when an object moves between consecutive views and that the local elements comprising a single object are close to one another. [10]

We found that, against our anticipation, SD was weaker in the same-absolute-location condition than in the different-absolute-location condition (Fig. 3d). We conjecture that the reduced SD in the former is likely to do with the repulsive sequential effect due to sensory adaptation [35], which can counteract SD. Our conjecture is in line with a recent volume of work supporting the view that 'attractive' and 'repulsive' sequential effects originate from distinct sources yet coexist [24,25,33,36–39]. Especially the fact that the repulsive sequential effect is pronounced when successive elements are matched in retinotopic coordinates [25,37] is consistent with our finding.

Our findings should be distinguished from debates about whether the spatial tuning of SD is governed by the distance in retinotopic coordinates [7] or in world-centered coordinates [5]. In our terms, both types of distance are considered as the distance in absolute coordinates. Instead, our findings suggest that SD is regulated by the distance between consecutive targets in object-centered coordinates. The object-centered space is similar to the world-centered space in that both use a reference frame outside the body, unlike the ego-centric retinotopic space. However, the world-centered space's reference frame is a fixed point that remains constant over time in the world, whereas the object-centered space's reference frame is an object within a group or a part of an object [14–17]. Given that the brain initially processes spatial information in the ego-centric retinotopic space, representing spatial information in the object-centered

space requires extra calculations to relocate elements within a new reference frame. Such extra calculations are known to occur in high-level associative cortical regions, including the supplemental eye field [18], parietal cortex [15], and lateral occipital-temporal cortex [14]. Our work indicates that the SD between successive local features is controlled in the object-centered space, suggesting that the neural mechanism responsible for SD is located in those associative cortices, or at least entails the translation of spatial information from retinotopic to object-centered coordinates. This possibility seems consistent with the view that SD is more likely to emerge at a high-level stage than at a low-level retinotopic stage where the adaptation-based repulsive bias is dominant [33,39,40].

In summary, we found that the correspondence in the object-centered space augments SD between consecutive targets. This spatial specificity of SD in relative coordinates assists the system in stabilizing perception between the consecutive targets that have the same identity when spatially grouped elements move together or are positioned close together.

Methods

Participants

For Experiment 1, we recruited fifteen participants (11 females), aged 22-35 years (27 on average) with normal or corrected-to-normal vision, through offline and online postings on the internet community of Seoul National University. They participated in a 90-minute web-based experiment and were compensated 15,000 KRW. Participants accepted the experiment's consent form and received a link to the online experiment via e-mail. For Experiment 2, we recruited participants in the same way in Experiment 1 and did not include those who had already participated in Experiment 1 or those who were unsuitable for eye tracking (e.g., wearing thick glasses). As a result, twenty-four participants (12 females), aged 19-33 years (25.5 on average) with normal or corrected-to-normal vision, participated in Experiment 2. They participated in a 90-minute offline experiment with eye tracking and were compensated 25,000 KRW. All gave written informed consent before the experiment. Both experiments were conducted in compliance with the safety guidelines for human experimental research, as approved by the Institutional Review Board of Seoul National University (IRB No. 2108/001-012).

Stimuli

Visual stimuli were Gaussian-enveloped sinusoidal gratings (Gabor patches) with a Michelson contrast of 0.3125 (in Experiment 1) or 0.3 (in Experiment 2). To promote effective fixation, a bullseye with a cross-hair inside was used as a fixation marker [41].

In Experiment 1, which was conducted online, it was not possible to precisely control the size of stimuli in degrees of visual angle (dva). Instead, we controlled the physical size of stimuli on each observer's screen and instructed them to maintain a distance of 60 cm between the screen and their eyes. Thus, instead of using dva, we use physical units (cm) to describe the spatial parameters of stimuli. Prior to participating in the experiment, each observer adjusted the size of a gray rectangle object on the screen to match the actual size of a credit card. We calculated the resolution of the observer's screen based on the adjusted size of the rectangle and ensured that the stimuli sizes matched their intended sizes. As for the Gabor patches, the sinusoidal grating had a spatial frequency of 0.6 cycles per centimeter, while the Gaussian envelope had a standard deviation of 1.3 cm. The vertical distance between the centers of the two Gabor patches was fixed at 6 cm. The two patches were placed horizontally, 12 cm to the right of the fixation marker, while the midpoint of the two patches alternated vertically between -1 cm and -5 cm to the bottom of the fixation marker (Fig. 2b). To make noise masks, we first created a 2.6 cm radius disc patch consisting of 0.2 cm x

0.2 cm squares whose brightness values were sample randomly from 0 to 255, and then low-pass filtered the disc patch using a 2D Gaussian filter (a standard deviation of 0.2 cm). The font size used for the retro-cue arrow was 1 cm in terms of pixel measurements. To report the orientation of the target, observers rotated two gray dots (each with a 0.375 cm radius) along a thick circle with a 2.5 cm radius around the fixation marker. The fixation marker, retro-cue, and orientation-reporting dots appeared in fixed positions on the upper left corner of the screen (Fig. 2a). For each trial, the orientations of two gratings were randomly selected from the angles of 0, 10, 20, ..., or 170, and the phases of the gratings were also randomly selected from the angles of 0, 90, 180, or 270. The stimuli were generated using customized Javascript codes.

In Experiment 2, we used a 24-inch LCD monitor (LG 24MP58VQ) to display the stimuli. The general setting is similar to Experiment 1, with the exception of the following details. The Gabor patches had a grating with a spatial frequency of 0.588 cycles per dva and a Gaussian envelope with a standard deviation of 1.1 dva. The horizontal distance between the centers of the two Gabor patches was fixed at 7 dva. The two patches were placed vertically, 9 dva to the bottom of the fixation marker, while the midpoint of the two patches alternated horizontally between 0 dva and 7 dva to the right of the fixation marker (Fig. 3a). To make noise masks, we first created a 2.6 cm radius disc patch consisting of 0.2 cm x 0.2 cm squares whose brightness values were sample randomly from 0 to 255, and then low-pass filtered the disc patch using a 2D Gaussian filter (a standard deviation of 0.2 cm). To make noise masks, we first low-pass filtered a pixel-by-pixel white noise pattern using a 2D Gaussian filter (a standard deviation of 0.2 dva) and applied a spatial 2D Gaussian filter (a standard deviation of 1.1 dva) to the low-pass filtered noise pattern. The noise masks' contrasts were normalized to a Michelson contrast of 0.9. The font size used for the retro-cue arrow was 1 dva in terms of pixel measurements. To report the orientation of the target, observers rotated two gray dots (each with a 0.5 dva) along a thick circle with a 2.5 dva around the fixation marker. The fixation marker, retro-cue, and orientation-reporting dots appeared in fixed positions on the upper center region of the screen (Fig. 3a). For each trial, the orientations of two gratings were randomly selected from the angles of 0, 1, 2, ..., 177, 178, or 179, and the phases of the gratings were also randomly selected from the angles from 0 to 360. The stimuli were generated using PsychToolbox for MATLAB [42].

Task

Experiment 1. In Experiment 1, observers performed 8 blocks of trials, with each block consisting of 50 trials. At the start of every block, observers were allowed to take a rest and reminded to keep their eyes 60 cm away from the monitor while fixating on the fixation marker throughout the block. Each trial was initiated by the appearance of the fixation marker on the upper-left side of the screen. After a delay of 700

ms, two gratings were shown for 500 ms, followed by noise patches lasting 800 ms. Observers had to remember the orientations of both gratings for a delay of 1 sec before one of the two gratings was indicated as a target with the retro-cue. The retro-cue, which was an upward or downward arrow lasting 800 ms, indicated the location of the target whose orientation should be reported. Once the retro-cue disappeared, observers reported the orientation of the retro-cued target grating by rotating the reporting dots around the fixation marker. They used two keys on a keyboard to rotate the reporting dots and had to complete their report within a 10-s time limit. Once observers confirmed their report, the reporting dots vanished, and the fixation marker in the next trial appeared after a delay of 500 ms. As a result, there was an interval of 1200 ms between the end of the previous trial and the stimulus onset of the next trial. Observers practiced sufficiently to understand the task before participating in the main experiment. As previously described, the two gratings alternatingly shifted their positions between consecutive trials. Specifically, their locations relative to the fixation marker defined in Cartesian coordinates were (12cm, 2cm) and (12cm, -4cm) in odd-number trials and (12cm, -2cm) and (12cm, -8cm) in even-number trials. In order to balance the number of trials between the same-relative-location and different-relative-location conditions, while also keeping the absolute distance between two consecutive targets within each condition, we modified the transition probability of retro-cue between consecutive trials (Fig. 5). As a result, the absolute distance between two consecutive targets was fixed at 4 cm in the same-relative-location and 2 cm in the different-relative-location condition. The number of trials was 2962 in the same-relative-location and 2918 trials in the different-relative-location condition.

Experiment 2. In Experiment 2, observers performed 9 blocks of trials, with each block consisting of 45 trials. The task structure was similar to Experiment 1, with the exception of the following aspects. The duration between fixation-marker onset and stimulus onset was 800 ms. The delay period between mask offset and retro-cue onset was 1500 ms. The duration of the retro-cue was 500 ms. To help ease the stress of observers who had to maintain their gaze and avoid blinking during the presentation of the fixation marker, we made the fixation marker disappear immediately after the noise patches disappeared. The midpoint of the two Gabor patches was randomized across trials, being either (0 dva, -9 dva) or (7dva, -9dva) relative to the fixation marker defined in Cartesian coordinates. Importantly, observers' gaze was monitored during stimulus presentation with an eye-tracker. When an observer's gaze significantly deviated from the fixation marker (see the "Eye movement tracking" below for details), the "Your eyes have moved" message was shown for 1 s below the fixation marker to alert the observer. When participants failed to fixate their gaze for 3 trials in a row, we instructed observers to pay more attention to maintain their gaze during stimulus presentation and re-calibrated the eye-tracker.

Eye movement tracking

We use a desktop-mounted Eyelink 1000 system (SR research, Ottawa, Canada) to track pupil diameter and gaze position at a sampling rate of 1000 Hz. To ensure accuracy, we calibrated the eye-tracker using a built-in 5-point routine (HV5) at the beginning of each block of trials and re-calibrated it if observers' fixation failed for three consecutive trials. We defined fixation failure with the following procedure. To determine the baseline gaze position for each trial, we calculated the average of the gaze position samples during the 240 ms before stimulus onset. If any gaze position sample deviated by more than 1.5 dva from this baseline, the trial was considered a fixation failure. We confirmed that the gaze deviations were quite small and did not vary depending on where the gratings were located (Fig. S3). Additionally, trials where eye blinks occurred during the fixation or stimulus presentation periods were also classified as fixation failure.

Data analysis

Discarding trials. In Experiment 1 and 2, we detected outlier trials with large errors and discarded them from further analysis. To identify these outliers, we used Tukey's method at an individual level by excluding trials in which estimation errors were less than the lower quartile (Q1) minus 1.5 times the interquartile range (IQR) or more than the upper quartile (Q3) plus 1.5 times the IQR. As a result, 5.08 % and 5.15 % of total trials were discarded for Experiment 1 and 2, respectively. In Experiment 2, we also discarded the fixation failure trials (2.75 %). We also identified trials where the distance between consecutive targets in absolute coordinates was 14 dva, which accounted for 12.25 % of total trials, did not include those trials in the main data analysis (as shown in Fig. 3c,d).

Data pooling across observers. Compared to other studies where only one target was remembered, orientation estimates tended to be noisier in our study, where two targets had to be remembered. There are several reasons for this, including an increase in memory load [43], contextual interferences from nearby stimuli [35,44], and pronounced stimulus-specific biases [45]. Due to the high degree of estimation noise, it is not reliable to fit the SD curve to individual data for each condition. Therefore, we pooled the data across individuals after demeaning the errors within individuals for statistical tests and visualization of SD, following previous studies [9,24].

Curve fitting. We fitted the derivative-of-Gaussian function, $y = xawce^{-(wx)^2}$, to the pooled data for each condition, where y is the estimation error; x is the previous target's orientation relative to the current target's orientation; c is a constant $\frac{\sqrt{2}}{e^{-0.5}}$; a and w are the free parameters associated with curve amplitude and width, respectively [5]. The best-fit parameters were determined by finding the values a and w that minimize the sum of squared errors with the BFGS algorithm. To avoid overfitting and promote fast saturation, we bounded the values of a and w within [-50, 50] and [0.02, 0.2], respectively. The best-fit value of a was used as the quantity representing how much the current target's orientation is attracted to the previous target's orientation. We estimated the reliability of the best-fit parameter value by defining 95% confidence intervals using the following non-parametric bootstrap procedure. First, the trials in a given condition were randomly resampled with replacements from each individual's data. Second, the resampled data were pooled across individuals. Third, the best-fit value of a was estimated by fitting the derivative-of-Gaussian function to the pooled data for each condition. We repeated this procedure 5,000 times to obtain the distribution of best-fit values and defined the 95% confidence interval based on this distribution.

Statistical test. To evaluate the significance of the difference in SD amplitude between any given two conditions, we carried out a permutation test [9,24] as follows. First, the labels of the two conditions (e.g., same-relative-location and different-relative-location conditions) were shuffled for each individual data. Then, the individual data were pooled across individuals according to the shuffled condition labels. Next, the best-fit value of a was estimated by fitting the derivative-of-Gaussian function to the pooled data for each of the two shuffled-label conditions. Lastly, we calculated the difference in the best-fit value of a between the two conditions. We repeated this procedure 10,000 times to obtain the distribution of these 'permutated' differences and calculated the p-value of the observed (non-shuffled) difference in SD amplitude based on this distribution.

Reference

1. Dong, D. W. & Atick, J. J. Statistics of natural time-varying images. *Network: Computation in Neural Systems* **6**, 345 (1995).
2. van Bergen, R. S. & Jehee, J. F. M. Probabilistic Representation in Human Visual Cortex Reflects Uncertainty in Serial Decisions. *J. Neurosci.* **39**, 8164–8176 (2019).
3. Felsen, G., Touryan, J. & Dan, Y. Contextual modulation of orientation tuning contributes to efficient processing of natural stimuli. *Network: Computation in Neural Systems* **16**, 139–149 (2005).
4. Cicchini, G. M., Mikellidou, K. & Burr, D. C. The functional role of serial dependence. *Proc. R. Soc. B.* **285**, 20181722 (2018).
5. Fischer, J. & Whitney, D. Serial dependence in visual perception. *Nat Neurosci* **17**, 738–743 (2014).
6. Kiyonaga, A., Scimeca, J. M., Bliss, D. P. & Whitney, D. Serial Dependence across Perception, Attention, and Memory. *Trends in Cognitive Sciences* **21**, 493–497 (2017).
7. Collins, T. The perceptual continuity field is retinotopic. *Sci Rep* **9**, 18841 (2019).
8. Fischer, C. *et al.* Context information supports serial dependence of multiple visual objects across memory episodes. *Nat Commun* **11**, 1932 (2020).
9. Pascucci, D. *et al.* Serial dependence in visual perception: A review. *Journal of Vision* **23**, 9 (2023).
10. Collins, T. Serial dependence occurs at the level of both features and integrated object representations. *Journal of Experimental Psychology: General* **151**, 1821–1832 (2022).
11. Liberman, A., Zhang, K. & Whitney, D. Serial dependence promotes object stability during occlusion. *Journal of Vision* **16**, 16 (2016).
12. Liberman, A., Manassi, M. & Whitney, D. Serial dependence promotes the stability of perceived emotional expression depending on face similarity. *Atten Percept Psychophys* **80**, 1461–1473 (2018).
13. Manassi, M. & Whitney, D. Illusion of visual stability through active perceptual serial dependence. *Sci. Adv.* **8**, eabk2480 (2022).
14. Committeri, G. *et al.* Reference Frames for Spatial Cognition: Different Brain Areas are Involved in Viewer-, Object-, and Landmark-Centered Judgments About Object Location. *Journal of Cognitive Neuroscience* **16**, 1517–1535 (2004).
15. Crowe, D. A., Averbeck, B. B. & Chafee, M. V. Neural Ensemble Decoding Reveals a Correlate of Viewer- to Object-Centered Spatial Transformation in Monkey Parietal Cortex. *Journal of Neuroscience* **28**, 5218–5228 (2008).
16. Driver, J. & Halligan, P. W. Can Visual Neglect Operate in Object-centred Co-ordinates? An Affirmative Single-case Study. *Cognitive Neuropsychology* **8**, 475–496 (1991).
17. Humphreys, G. W., Gillebert, C. R., Chechlacz, M. & Riddoch, M. J. Reference frames in visual selection: Reference frames in visual selection. *Ann. N.Y. Acad. Sci.* **1296**, 75–87 (2013).
18. Olson, C. R. Brain representation of object-centered space in monkeys and humans. *Annual review of neuroscience* **26**, 331–354 (2003).
19. Kahneman, D., Treisman, A. & Gibbs, B. J. The reviewing of object files: Object-specific integration of information. *Cognitive psychology* **24**, 175–219 (1992).
20. Boi, M., Vergeer, M., Ogmen, H. & Herzog, M. H. Nonretinotopic Exogenous Attention. *Current Biology* **21**, 1732–1737 (2011).
21. Wutz, A., Drewes, J. & Melcher, D. Nonretinotopic perception of orientation: Temporal integration of basic features operates in object-based coordinates. *Journal of Vision* **16**, 3 (2016).

22. Boi, M., Ogmen, H., Krummenacher, J., Otto, T. U. & Herzog, M. H. A (fascinating) litmus test for human retino- vs.non-retinotopic processing. *Journal of Vision* **9**, 5–5 (2009).
23. Corbett, J. E., Fischer, J. & Whitney, D. Facilitating Stable Representations: Serial Dependence in Vision. *PLoS ONE* **6**, e16701 (2011).
24. Fritsche, M., Mostert, P. & de Lange, F. P. Opposite Effects of Recent History on Perception and Decision. *Current Biology* **27**, 590–595 (2017).
25. Fritsche, M., Spaak, E. & de Lange, F. P. A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *eLife* **9**, e55389 (2020).
26. Boi, M., Ogmen, H. & Herzog, M. H. Motion and tilt aftereffects occur largely in retinal, not in object, coordinates in the Ternus-Pikler display. *Journal of Vision* **11**, 7–7 (2011).
27. Corbett, J. E. The Whole Warps the Sum of Its Parts: Gestalt-Defined-Group Mean Size Biases Memory for Individual Objects. *Psychol Sci* **28**, 12–22 (2017).
28. Kubovy, M., Holcombe, A. O. & Wagemans, J. On the Lawfulness of Grouping by Proximity. *Cognitive Psychology* **35**, 71–98 (1998).
29. Wagemans, J. *et al.* A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin* **138**, 1172–1217 (2012).
30. Manassi, M., Kristjánsson, Á. & Whitney, D. Serial dependence in a simulated clinical visual search task. *Sci Rep* **9**, 19937 (2019).
31. Taubert, J., Alais, D. & Burr, D. Different coding strategies for the perception of stable and changeable facial attributes. *Sci Rep* **6**, 32239 (2016).
32. Huang, J. & Sekuler, R. Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision* **10**, 24–24 (2010).
33. Pascucci, D. *et al.* Laws of concatenated perception: Vision goes for novelty, decisions for perseverance. *PLoS Biol* **17**, e3000144 (2019).
34. Barbosa, J. & Compte, A. Build-up of serial dependence in color working memory. *Sci Rep* **10**, 10959 (2020).
35. Gibson, J. J. Adaptation, after-effect, and contrast in the perception of tilted lines. II. Simultaneous contrast and the areal restriction of the after-effect. *Journal of Experimental Psychology* **20**, 553–569 (1937).
36. Moon, J. & Kwon, O.-S. Attractive and repulsive effects of sensory history concurrently shape visual perception. *BMC Biol* **20**, 247 (2022).
37. Pascucci, D. & Plomp, G. Serial dependence and representational momentum in single-trial perceptual decisions. *Sci Rep* **11**, 9910 (2021).
38. Sadil, P., Rosemary Cowell & Huber, D. E. *The Push-pull of Serial Dependence Effects: Attraction to the Prior Response and Repulsion from the Prior Stimulus*. https://osf.io/f52yz (2021) doi:10.31234/osf.io/f52yz.
39. Sheehan, T. C. & Serences, J. T. Attractive serial dependence overcomes repulsive neuronal adaptation. *PLoS Biol* **20**, e3001711 (2022).
40. Schwiedrzik, C. M. *et al.* Untangling Perceptual Memory: Hysteresis and Adaptation Map into Separate Cortical Networks. *Cerebral Cortex* **24**, 1152–1164 (2014).
41. Thaler, L., Schütz, A. C., Goodale, M. A. & Gegenfurtner, K. R. What is the best fixation target? The effect of target shape on stability of fixational eye movements. *Vision Research* **76**, 31–42 (2013).
42. Kleiner, M., Brainard, D. & Pelli, D. What's new in Psychtoolbox-3? (2007).

43. Bays, P. M., Catalao, R. F. G. & Husain, M. The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision* **9**, 7–7 (2009).
44. Cicchini, G. M., Benedetto, A. & Burr, D. C. Perceptual history propagates down to early levels of sensory analysis. *Current Biology* **31**, 1245-1250.e2 (2021).
45. Balikou, P. *et al.* Independent sources of anisotropy in visual orientation representation: a visual and a cognitive oblique effect. *Exp Brain Res* **233**, 3097–3108 (2015).

Acknowledgements

This research was supported by the Research Grant from Seoul National University (339-20220013), and by the Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (No. NRF-2021R1F1A1052020).

Author contributions

J.L. and S.L. designed the experiments, analyzed results, and wrote the manuscript. J.L. collected data.

Competing Interests Statement

The authors declare no competing interests.

Figures

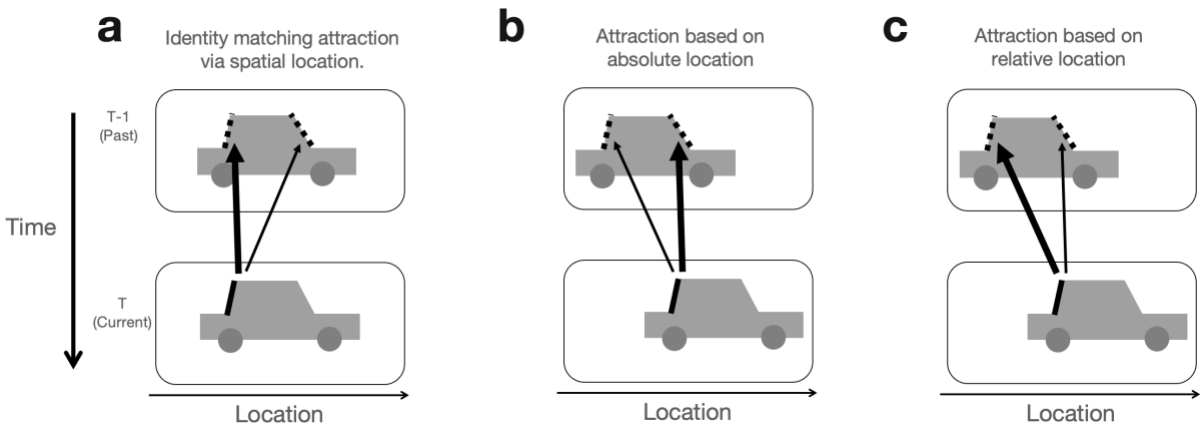

Figure 1. Identity-matching problems in serial dependence (SD) under static (**a**) versus dynamic (**b,c**) viewing conditions. Spatial correspondence can be an essential cue to identifying which elements match in identity between consecutive views. Thus, SD is expected to be transmitted preferentially between the consecutive elements with spatial correspondence. There are two types of spatial correspondence, one defined in absolute coordinates and the other in relative coordinates. These two types are indistinguishable when an object remains static across views (**a**). For example, the two consecutive elements matched in identity (the rear windshield, as indicated by the thick arrow in (**a**) correspond not only in absolute coordinates but also in relative coordinates. By contrast, the absolute and relative spatial correspondences can be distinguished when an object shifts in position across views (**b,c**). In such a dynamic viewing condition, it is not the absolute (**b**) but relative (**c**) spatial correspondence that determines whether a given pair of consecutive elements match in identity. Thus, if SD is transmitted between consecutive elements with the same identity, it is expected to be more pronounced between the elements corresponding in relative coordinates than those in absolute coordinates. The thickness of the arrows represents the magnitude of SD.

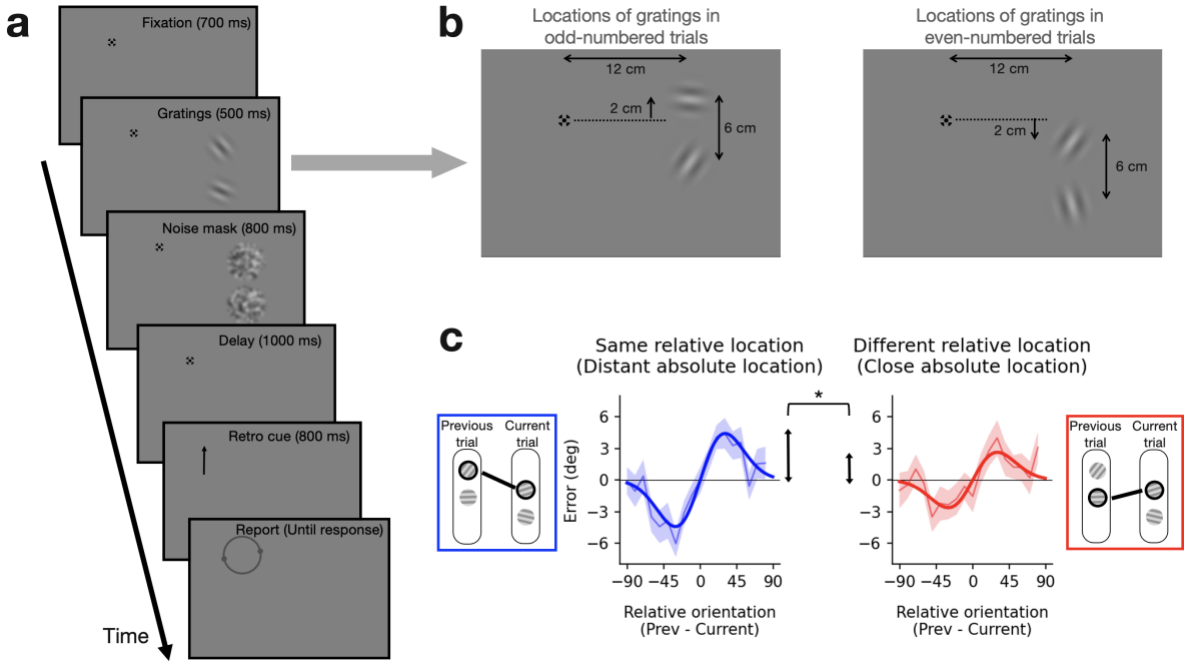

Figure 2. Design and results of Experiment 1. **a,** Trial structure. While fixating on the fixation cross, observers briefly viewed two Gabor patches in the right visual field, which were followed by masks. After a short delay, they reported the orientation of the patch retro-cued by an arrow. An inter-trial interval of 1.3 s separated trials, and the fixation cross pre-warned the trial onset**. b,** Spatial arrangement of stimuli. On any given trial, two vertically arranged Gabor patches appeared simultaneously. The distance between the two patches remained fixed while their positions alternated vertically between trials. **c,** Comparison of SD between the same-relative-location (left, blue) and different-relative-location (right, red) conditions. The insets within the colored boxes show example sequences of Gabor patches in consecutive trials. The gratings marked by the black circles and lines indicate the retro-cued targets. The two panels in the middle plot the observers' orientation estimation errors in the current trial as a function of the difference in target orientation between the previous and current trials. The thin colored lines and shades represent the across-observer averages of the errors and their standard error of means. The thick curves with color represent the Gaussian derivatives fitted to the observed errors, quantifying the magnitude of SD with amplitudes, as indicated by the vertical arrows. The star between the two panels indicates that SD differed in amplitude between the same-relative-location and different-relative-location conditions ($p<0.05$).

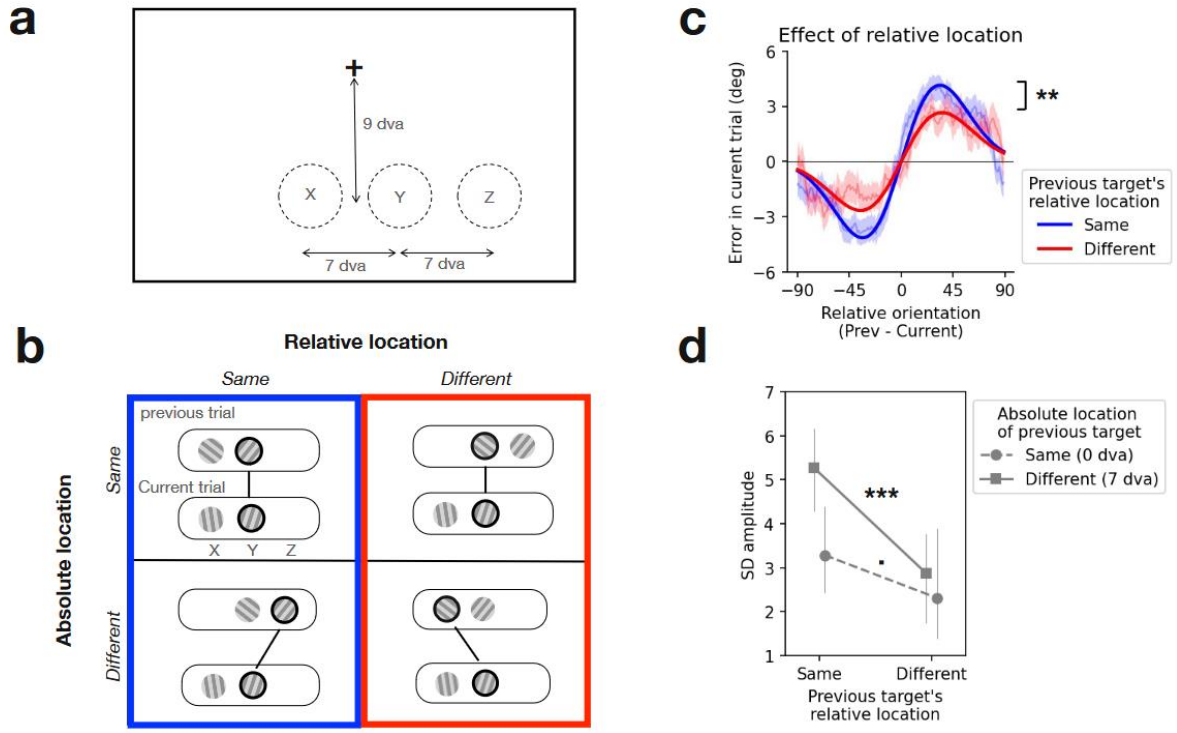

Figure 3. Design and results of Experiment 2. **a**, Spatial arrangement of stimuli. On any given trial, two horizontally arranged Gabor patches appeared randomly in two adjacent positions out of three ('X and Y' or 'Y and Z') below the fixation cross. **b**, Spatial correspondences of the retro-cued targets between consecutive trials. Each panel shows an example sequence of displays in consecutive trials (top and bottom round rectangles for the previous and current trials, respectively), where the black line and circles indicate the retro-cued targets. All trials can be sorted into four target-location conditions depending on whether the two consecutive targets match in absolute and relative coordinates: (i) matched in both coordinates (top left); (ii) matched only in absolute coordinates (bottom left); (iii) matched only in relative coordinates (top right); (iv) matched in neither coordinates (bottom right). **c,** Comparison of SD between the same-relative-location (blue) and different-relative-location (red) conditions. The observers' orientation estimation errors in the current trial are plotted as a function of the difference in target orientation between the previous and current trials. The thin colored lines and shades represent the across-observer averages of the errors and their standard error of means. The thick curves with color represent the Gaussian derivatives fitted to the observed errors, quantifying the magnitude of SD with amplitudes. The blue and red curves derived from the trials of the conditions marked by the corresponding colors in B. The two stars indicate that SD differed in amplitude between the same-relative-location and different-relative-location conditions ($p<0.01$). **d**, Comparison of SD in magnitude across the target location conditions. The across-observer mean amplitudes of SD are plotted against the location conditions defined in relative coordinates, separately for the two

location conditions defined in absolute coordinates. Error bars are 95% confidence intervals of SD amplitude determined by a bootstrap procedure. The stars and dot indicate the significance of the decrease in SD (. $p<0.1$; ** $p<0.01$; one-sided permutation test).

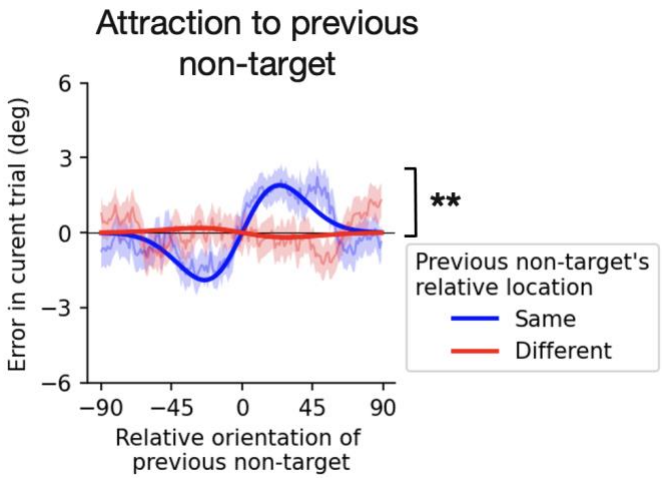

Figure 4. Attractive bias toward previous non-target orientation in Experiment 2. The observers' orientation estimation errors in the current trial are plotted as a function of the difference in orientation between the non-target in the previous trial and the target in the current trials, separately for when the non-target and target matched in relative coordinates (blue) and when not (red). The thin colored lines and shades represent the across-observer averages of the errors and their standard error of means. The thick curves with color represent the Gaussian derivatives fitted to the observed errors, quantifying the magnitude of SD with amplitudes. The two stars indicate that SD differed in amplitude between the same-relative-location and different-relative-location conditions ($p<0.01$).

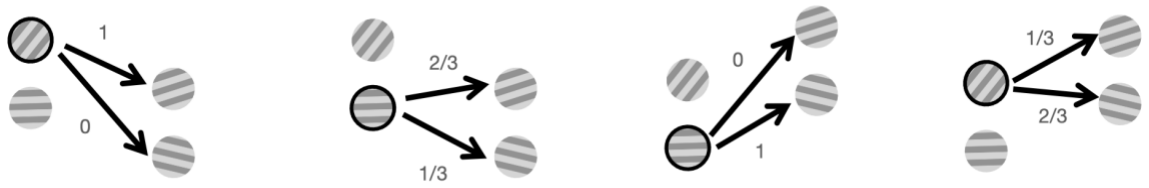

Figure 5. Transition probability of target location between consecutive trials in Experiment 1. Each panel shows the transition probabilities associated with the target at a specific location in the previous trial. The retro-cued targets in the previous trial are marked by black circles in the left column. The arrows indicate the targets in the current trial, the numbers on which are the transition probabilities.