Full Title: Learning efficient representations of heterogeneity in attractors for working memory

Short Title: Parametric learning in working memory

Tahra L Eissa[1,*] and Zachary P Kilpatrick[1,2]

[1]Department of Applied Mathematics, University of Colorado Boulder, Boulder, CO, USA.

[2]Institute of Cognitive Science, University of Colorado Boulder, Boulder, CO, USA.

[*]corresponding author: tahra.eissa@colorado.edu

Abstract

Natural stimulus distributions often exhibit biases that we may learn to predict what we will observe next. Correspondingly, humans performing delayed estimation tasks are biased in favor of more commonly represented stimuli (e.g., common colors in visual tasks). These systematic biases may emerge from attractors in neural ensemble activity representing memoranda during delay periods. However, it is unclear how these heterogeneous representations are learned and how they may be represented in the brain. Here, we present a recurrently coupled neural network model that can learn heterogeneous input distributions via a combination of long term potentiation and homeostatic plasticity, generating an underlying potential landscape that governs the preferred positions of a bump attractor on a ring. The resulting spatially-heterogeneous synaptic network probabilistically represents the experienced stimulus history of the environment. We validate this theory using recently published response data from a delayed estimation task in which the distribution of stimuli was heterogeneous. Our results suggest that models that learn heterogeneities in the stimulus distribution explain the data more often than models for which connectivity is uniform throughout the duration of the experiment.

Introduction

Working memory is a system that maintains and computes information over short timescales without a constant source of input and has been identified as a core feature of cognition. Knowing the limitations of this system can help determine the structure of problems people can use it to solve [1]. Tasks can quantify working memory limits and biases, for instance, by measuring the statistics of errors made when recalling stimulus features (e.g., position, orientation, frequency) after a delay [2]. In visual working memory tasks, humans are asked to recall the features of an object (or objects), such as location, color, or shape, a short time later [1, 3–5]. When these features lie on a continuum, responses do as well, giving

finely resolved measurements of the direction and magnitude of errors on each trials [6, 7]. One common observation in such continuous report tasks is that error magnitudes increase roughly linearly with time as variance of a diffusion process would [8, 9]. Such psychophysical observations can help guide the development of neural theories for working memory limitations.

One well-validated and predictive theory of how the brain encodes information over short periods of time associates remembered stimulus values with persistent neural activity sustained in recurrently coupled excitatory neurons tuned to feature values [10]. Broadly tuned inhibitory neurons driven by the excitatory population then stabilize this activity into a localized (in feature space) structure called an activity *bump* [11, 12]. Heterogeneity and variability in neural tuning and synaptic connectivity can cause this bump of activity to wander, leading to observable trial-by-trail errors and biases [13–15]. It has even been suggested that such biases may be advantageous, if they reflect the structure of the environment and sequences of stimuli might expect to see [16, 17]. Yet, the mechanisms by which the brain learns these biases is not well described.

Systematic biases in working memory have been observed to take a number of quantifiable forms. First, working memory may exhibit serial bias, whereby stimulus values on the previous trials may attract or repel the retained memory of the stimulus value on the current trial [18]. Analogous attractive biases also emerge when subjects retain the values of multiple stimuli within a single trial [19]. Alternatively, subjects may exhibit biases to longer term trends, such as an expected distribution of stimulus value. This can include preferences for focal colors [15, 20], orientations [21] and cardinal directions [22]. Since features like cardinal direction or colors are often overrepresented in the natural world [23, 24], these sustained biases could be the result of experience-dependent learning based on their natural heterogeneous representation in our environments. One candidate for learning such systematic heterogeneity in neural networks is spatial heterogeneity in the distribution of synapses from stimulus-tuned excitatory neurons. Synaptic heterogeneity has been shown to reduce error by maintaining representations that are less susceptible to fluctuations [25–27]. Taking this one step further, spatial heterogeneity could be chosen to responses so they are distributed similarly to the expected distribution of stimulus values. Common features would be recalled with less error, improving cognitive efficiency [28].

Here, we propose that overexposure to common features leads to experience-dependent learning of the environmental statistics, which modulates synaptic connectivity to produce representation biases. We modeled the effects of heterogeneous environments on working memory dynamics in mechanistic (neural field) models and their low-dimensional reductions (particles), considering static and learning models. We established a plausible neuromechanistic learning rule for the modulation of synaptic weights based on experience, via long-term potentiation and homestatic plasticity. We compared these results to human behavior and found that most subjects exhibit strategies that are best described by learning models, supporting the hypothesis that long-term representation biases are indeed the result of experience-dependent learning.

Results

We began with the premise that the environment includes heterogeneous distributions of parameters, and particular stimulus values (e.g., locations or colors) are overrepresented in nature and thus, statistically more likely to occur (Fig. **1a**). We considered delayed response/estimation tasks in which subject either had to report a feature value of a remembered stimulus, whose values are defined on a ring (e.g., one feature dimension that wraps periodically). For illustrative purposes, we will focus on examples in which subjects must recall colors, though equivalent results can be produced for

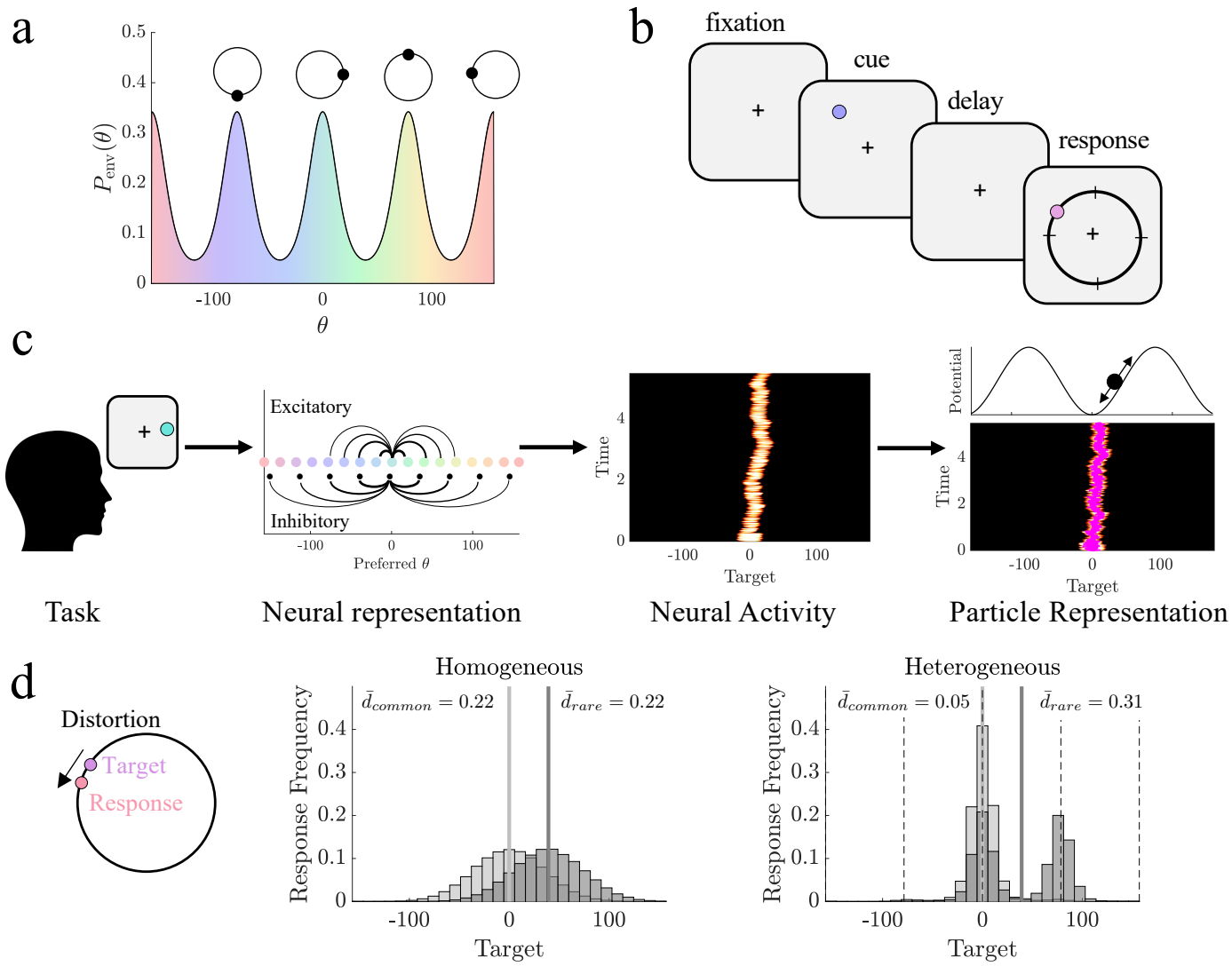

Figure 1: Environmental heterogeneity reflected in working memory heterogeneity. **a.** Environments with feature heterogeneity, such as orientations at cardinal directions being more common and overrepresentation of certain colors, produce heterogeneous probability distributions with peaks at the overrepresented features. **b.** Schematic of the delayed estimation task, which requires subjects to remember a feature of the cue (e.g., color) and report it following some short delay period. **c.** The target is represented in the brain by activation of a subpopulation of feature-specific excitatory neurons with local recurrent connectivity and broad inhibition. The target representation retained as a bump of sustained neural activity during the delay period. The dynamics of this bump can be reduced to that of a particle, which stochastically wanders over time. Given spatial heterogeneity in synaptic connectivity, the particle will inherit a nontrivial energy landscape with basins that represent attractors in the system. **d.** Distortion computed as the circular distance between the target and responses is influenced by the heterogeneity of the system. In homogeneous networks with no bias, response trends at common environmental targets ($\theta = 0$, light grey) and rare targets ($\theta = \pi/4$, dark grey) are equivalent, giving the same mean distortion ($\bar{d}$). Given heterogeneity aligned so that response biases match the environmental distribution of stimulus feature values, errors are reduced in the vicinity of common stimulus feature values (dashed lines). Parameters used:$\sigma = 0.05$, $T = 5$, $n = 0, 4$, $A = 1$, $N_{\text{Sims}} = 1e5$.

orientation and location. Therefore, we described the probability distribution of observing a particular feature value θ as

$$P_{env}(\theta) = e^{A\cos(m\theta)},$$

where A describes the amplitude and m describes the number of peaks in the probability distribution. Unless otherwise stated, all subsequent results assume that $m = 4$, so peaks are centered at cardinal angles for all heterogeneous input distributions. We considered models based on the delayed response/estimation task [29, 30], in which an observer is shown a number of items and probed about stimulus feature values (e.g., location, orientation, or color) after some short delay period (Fig. **1b**) to theorize how environmental heterogeneities might determine how memoranda are retained.

Our models are motivated by previous observations showing human performance on delayed estimation tasks degrades over time such that response variance increases roughly linearly (suggesting an underlying diffusive process driving the forgetting process) [8]. Such diffusive degradation of a remembered stimulus featured estimated has been modeled in neural networks with sustained and stimulus-tuned excitation along with broad stabilizing inhibition with neural and synaptic fluctuations that drive the stochastic wandering of a localized region of persistent activity (bump) [10, 31]. Excitatory neurons with similar feature preferences are strongly connected by recurrent connectivity, generating this self-sustained activity [11]. Since the wandering location (centroid) of the activity bump is a proxy for the remembered stimulus feature value [13, 14], it can be useful to consider low-dimensional models that simply describe diffusion of a particle along a potential function that describes any spatial bias generated by spatial heterogeneity in synaptic connectivity [25, 26, 32], which can be more easily fit to response data than full neural circuit models [13, 15, 17, 33]. The particle's movement represents that of the bump of neural activity and is shaped by the landscape of an energy potential that can be varied to incorporate bias (Fig. **1c**).

To quantify the amount of error between the target and response as performed in both psychophysical experiments and previous modeling efforts [14, 26], we computed the mean total distortion for all target and the response locations following the delay period:

$$\bar{d} = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} P(\theta_{env})P(\theta_{resp}|\theta_{env})(1-\cos(\theta_{resp}-\theta_{env}))d\theta_{env}d\theta_{resp}$$

in which the conditional probability density $P(\theta_{resp}|\theta_{env})$ describes the likelihood of a response θ_{resp} given an input value θ_{env}. For a particular target location θ_{env}, the mean total distortion $\bar{d}$ and response distributions are impacted by the heterogeneity in representation network. Results are generated from particle models whose energy landscapes are inherited from spatial synaptic heterogeneity of recurrent excitation, as can be derived asymptotically [32, 34]. For networks with spatially homogeneous synaptic connectivity, response distributions and mean distortion are similar for common or rare targets. In networks with spatial synaptic heterogeneity matched to the environment, distortion is reduced at common targets and enhanced at rare targets (Fig. **1D**). This effect is the result of an attractive bias toward the peaks of $P_{env}(\theta)$, so response errors are reduced when the target is at the peak, but responses are repelled from locations corresponding to rare feature values.

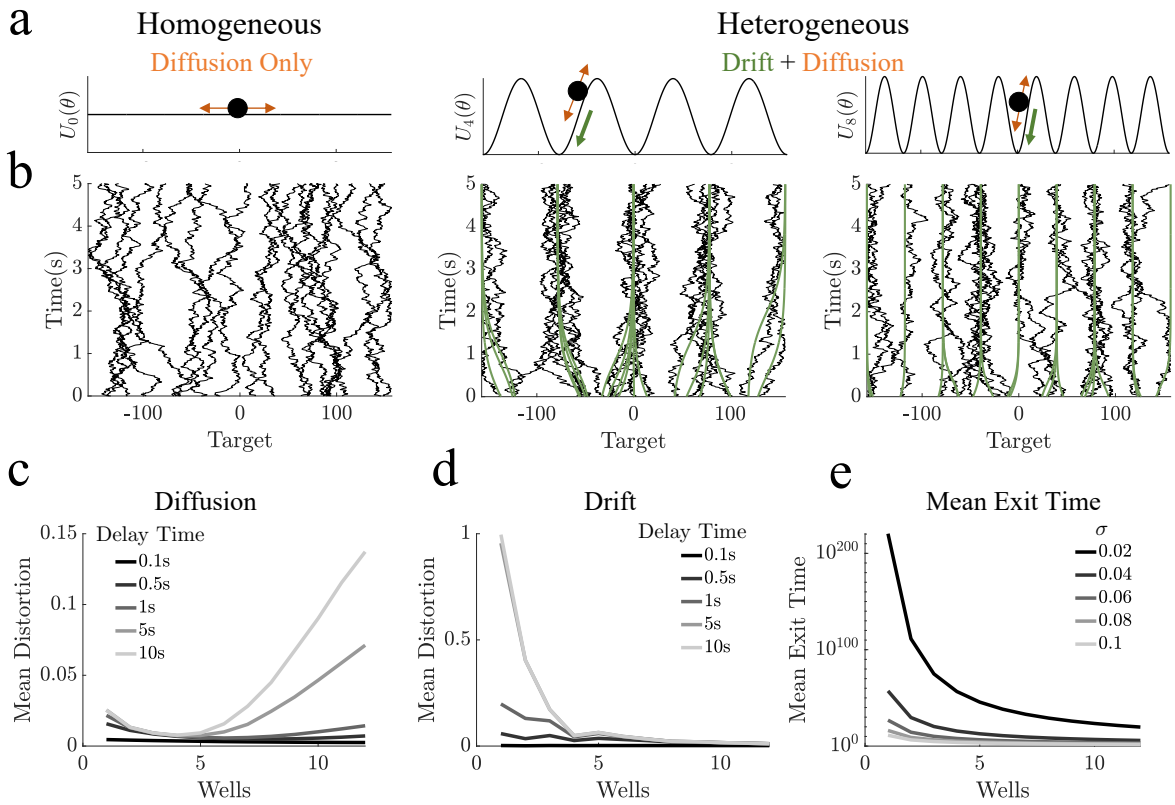

Figure 2: Particles in heterogeneous landscapes are drawn toward attractors. **a.** Schematics of the homogeneous particle model, with only diffusion, and heterogeneous particle models, with potential-driven drift and diffusion. Parameters used: $A = 1$, $n = 0, 4, 8$. **b.** Example of particle trajectories in the homogeneous and heterogeneous energy landscapes sampled from a uniform environmental distribution, corresponding with landscapes above in **a**. In the homogeneous landscape, diffusion incorporates noise into the system. In the heterogeneous landscapes, the inclusion of drift draws the particle toward the wells of attraction over time but diffusion can cause the particle to "jump" wells. Drift-only process shown in green for comparison. Parameters used:$\sigma = 0.05$, $T = 5$, $n = 0, 4, 8$, $A = 1$, $N_{\text{Trials}} = 30$. **c.** Mean distortion in a heterogeneous landscape with only diffusion (targets sampled at attractor point) and moderate diffusion ($\sigma = 0.05$). Parameters used: $A = 1$, $N_{\text{Sims}} = 1e5$. **d.** Mean distortion in a heterogeneous landscape with only drift. Parameters used: $A = 1$, $N_{\text{Sims}} = 1e5$. **e.** Mean exit time for a particle to leave the current well of attraction. Computed as in [26, 35].

We now demonstrate how the energy landscapes of particle models (reflecting spatial synaptic heterogeneity) impact response distributions. Combining this with the impacts of different distributions of true stimulus values, we can ultimately understand how expected and actual heterogeneity interact to determine error (distortion) in a subject's responses. Ultimately, this can guide us toward an understanding of how expected heterogeneity might be learned from the actual heterogeneity of the environment and how this might aid coding during delayed estimation.

Static Particle Models

Homogeneous Environments

To interpret the effects of heterogeneity on working memory, we considered a particle model in which the particle's potential landscape was modulated to represent heterogeneity (Fig. **2a**) described by

$$U(\theta) = -A_p \cos(n\theta)$$

where A_p described the amplitude and n described the number of wells or preferred locations assumed by the observer. Note, we could have chosen a more general form for $U(\theta)$, but the simplicity of this form allows us to probe how matching and mismatching the energy landscape to the true stimulus distribution can shape observer distortion. The movement of the particle in time can be described by the stochastic differential equation

$$d\theta(t) = -U'(\theta(t))dt + \sigma dW(t)$$

where the evolution of a particle is comprised of two parts: 1. descent of the particle toward the wells of attractions of the potential $U(\theta(t))$ (drift), and 2. noise (diffusion), given by $W(t)$, a Wiener process. Considering a particle model with a homogeneous energy landscapes ($A_p \equiv 0$), we see the memory of the original stimulus feature value during the delay period exhibits only noisy fluctuations. On the other hand, particles stochastically evolving in heterogeneous energy landscapes ($A_p > 0$) tended toward the periodically placed attractors at $\theta = \pm 2\pi j/n$ ($j = 0, ..., n-1$) (Fig. **2b**), reflecting a bias.

We then quantified the error made in environments with a homogeneous (flat) distribution of feature values. Previous studies have shown that heterogeneity can serve to reduce the corrupting effects of diffusion on delayed estimates [25,26, 36]. Distortion in the particle model was influenced by drift and diffusion in distinct ways. First, we fixed the diffusion coefficient and sampled responses based on stimulus feature values sampled exclusively at the observer's associated attractor points to assess the impact of diffusion alone in heterogeneous landscapes, and we found a non-monotonic relationship between the number of wells and distortion. This relationship was enhanced by longer delays (Fig. **2c**), due to the increasingly close proximity of nearby wells as the number increases, reducing the energy needed for the particles to 'jump' into other basins. Low (high) levels of diffusion showed reduced (enhanced) mean distortion (Fig. **S1**). In contrast, when removing noise so that particles in a heterogeneous landscape were only subject to drift (no noise), distortion decreased considerably as the number of attractors was increased (Fig. **2d**). Finally, we computed the mean exit time for a particle to leave a well (region in which particle flows toward attractor given no noise) as a measure of attractor number (Fig. **2e**) as in [26, 35]. We found that mean exit time decreased as we increased the number of wells and as well as diffusion, confirming that the particle diffuses more easily in landscapes with more wells.

Heterogeneous Environment

We then compared the mean distortion of the homogeneous and heterogeneous models when exposed feature values were distributed heterogeneously. Modulating the location and number of wells in the particle model altered the conditional probability $P(\theta_{Resp}|\theta_{env})$ and marginal probability $P(\theta_{resp})$ varying the position and height of peaks, reflecting both the presence of attractors in the energy landscape and more frequent feature values (Fig. **3a**). Matching the number of particle wells to the number of peaks in the input distribution, we found that distortion was significantly reduced at attractor (common) locations as compared to the homogeneous model, but showed comparable levels of distortion at oblique (uncommon) angles (Fig. **3b**; bootstrapped distortion, $p < 0.05$). We analyzed the effect of increasing the depth

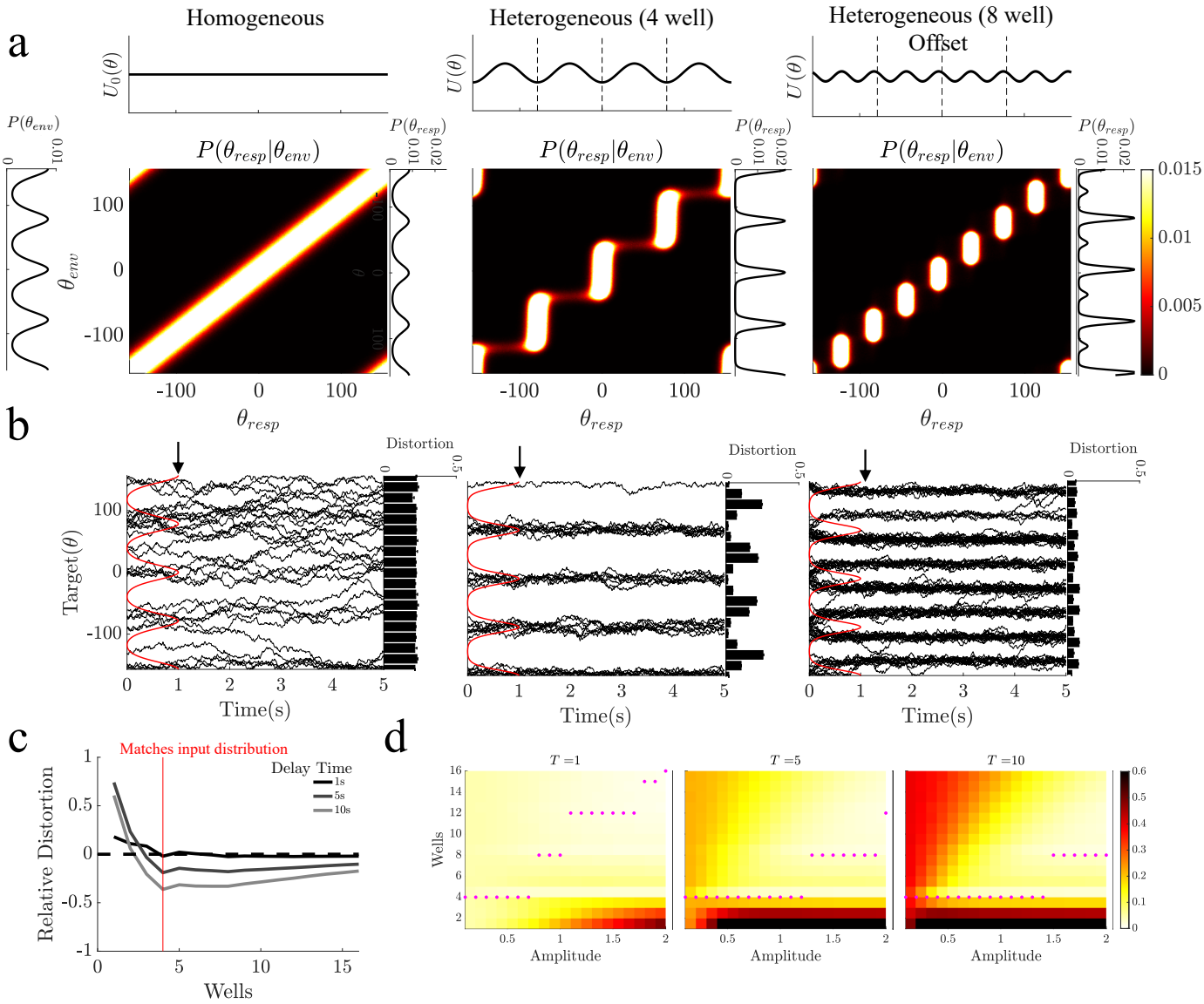

Figure 3: Distortion is reduced in heterogeneous environments when energy landscapes are heterogeneous and matched. **a.** Comparing energy landscapes ($U(\theta)$) and heterogeneous feature value distribution ($P(\theta_{env})$), we find the conditional probability of response $P(\theta_{Resp}|\theta_{env})$ and the marginal probability of response $P(\theta_{resp})$ for particle models with homogeneous and (two) heterogeneous (four wells at environmental distribution peaks and eight offset from the peak locations). Parameters used: $D = 0.05$, $T = 1$, $n = 0, 4, 8$, $A = 1$, $offset = \pi/4$, $N_{\text{Trials}} = 1e5$. **b.** Example trajectories of particle movement sampled from the heterogeneous environmental distribution (shown in red) and the particle landscapes displayed in **a**. Mean bootstrapped distortion for corresponding angles shown on right. Arrow denotes delay time at which conditional and marginal distributions in **a** were computed. Parameters used: $D = 0.05$, $T = 5$, $n = 0, 4$, $A = 1$, $N_{\text{Sims}} = 1e5$, $N_{\text{Boot}} = 1e3$. **c.** Relative distortion compared between the homogeneous and heterogeneous particle models (negative values denote reduced distortion when using a model with a heterogeneous energy landscape). Red vertical line denotes where the number (and position) of attractors in the particle model matches the number (and position) of peaks in the environmental distribution. Parameters used: $D = 0.05$,$A = 1$, $N_{\text{Sims}} = 1e5$. **d.** Mean distortion as amplitude and number of wells was varied for three example delay periods. Optimal particle model identified based on minimum mean distortion (magenta dots). Parameters used: $D = 0.05$,$A = 1$, $N_{\text{Sims}} = 1e5$.

of attractor wells and those that were offset from the environmental peaks, and found that when the number of wells was increased, the decrease in distortion was observed across all values of θ. However, the mean distortion was never reduced to the same level as when the positions of attractors was chosen to exclusively match the location of the most common feature values (aligning wells with peaks in the distribution).

We then identified the particle landscapes that best reduced distortion (on average) across all stimulus values θ for

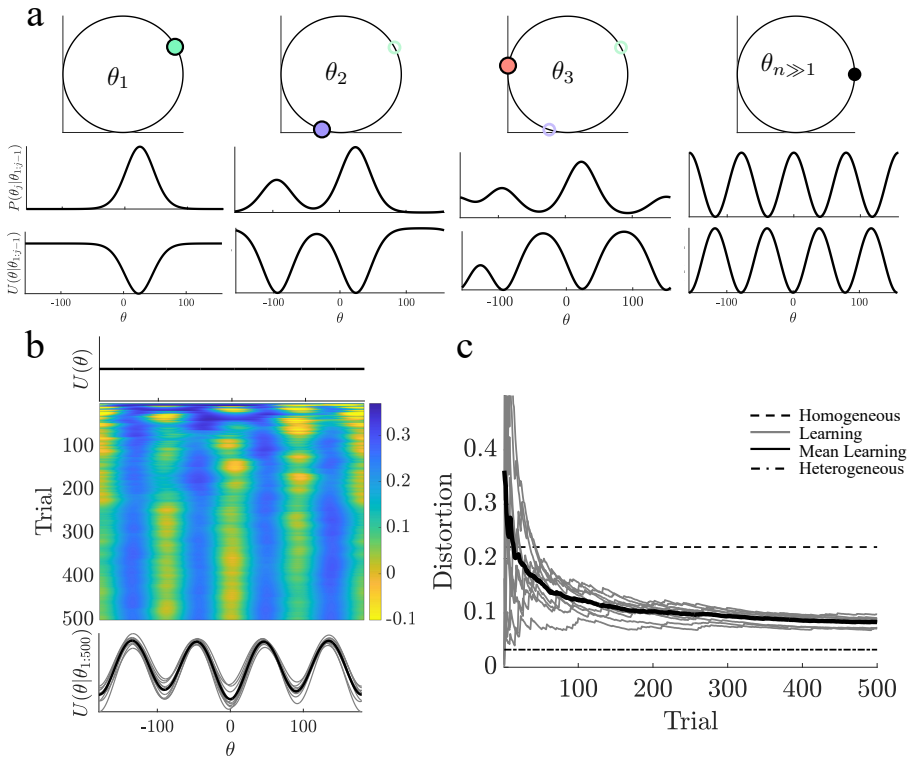

Figure 4: Experience-dependent learning leads to a reduction in distortion over time. **a** Schematic of experience-dependent learning in a particle model. Based on the target observed on each trial θ, the predicted distribution is updated for the subsequent trial $P(\theta_j|\theta_{1:j-1})$. Correspondingly, the particle landscape is updated to incorporate a well at the trial location, averaged across trials. Over the course of many trials, the predicted distribution recovers the environmental distribution and the particle landscape recovers the static heterogeneous landscape. **b** Heatmap showing landscape updating over the course of many trials, which recovers heterogeneity matching the static model. Top trace shows initial landscape. Bottom trace displays the landscape produced by the final trial. Grey traces are 10 examples of the learning model, black trace is the average learning model's landscape. Parameters used: $D = 0.05$, $T = 5$, $\kappa = 8$, $h = 0.25$, $s = 5$, $N_{\text{Trials}} = 500$. **c** Running average of the learning model's distortion. Homogeneous (heterogeneous) model's distortion shown as a dashed (dot dashed) line for reference. Grey traces are individual simulations, black trace shows simulation average. Parameters used: $D = 0$, $T = 5$, $\kappa = 8$, $h = 0.25$, $s = 5$, $N_{\text{Trials}} = 500$, $A = 1$, $n = 0, 4$.

a given environment. Varying number of wells and the delay period, we computed the relative distortion between the heterogeneous models and the homogeneous model and found that almost all heterogeneous models exhibited a reduction in distortion, with the largest reduction appearing when the number of wells matched the number of peaks in the input distribution ($m = n$, Fig. **3c**). To confirm if this solution was optimal, we then computed the mean distortion as both amplitude and the number of wells was varied and found the minimal distortion for a given delay period. Notably, for larger energy landscape amplitudes, distortion was not minimized by an equivalent number of wells to the input peaks, but instead corresponded to when there were more attractors than peaks in the environmental distribution, often at the harmonics of the input peaks (Fig. **3c**). This is consistent with previous results, which have shown that even when environmental distributions are homogeneous, it can be advantageous to introduce heterogeneous energy landscapes to slow the effective rate of diffusion during the delay [25, 26, 36]. These results were observed across different noise amplitude, though the distortion minimization by harmonics occurs at lower amplitudes when there was less noise (Fig. **S2**, which may be due to the stronger drift effects as compared to diffusion.

Experience-dependent learning particle models

We then extended our particle model to then consider an experience-dependent update rule that could learn hetergeneities matched to the heterogeneities in the environmental distribution. Such a rule is based on the assumption that subjects update their beliefs about the environment based on a sequential inference process. Thus, for each observation, the predicted probability distribution of the subsequent target is updated with a peak at the current trial's target location, merging with a running posterior. In the particle model, this update is inherited as an inverted Gaussian on a circle (Von Mises distribution), creating a well at the input location that is normalized to maintain the level of attraction (Fig. **4a**). Thus, for each observation, the estimated particle landscape $U_{est}(t)$ is updated on trial n as

$$U_{est}^{n}(\theta) = U_{est}^{n-1}(\theta) + \frac{s(h - e^{\kappa\cos(\theta-\theta_{inp})-1})}{n}$$

where s is the scaling (depth) of the well, h is a shift, and κ controls the spread (width) of the well. Over the course of many trials, heterogeneity of the input distribution was learned by the energy landscape, placing wells at peaks in the environmental distribution (Fig. **4a,b**). The mean distortion of the learning model decreased with experience, trending toward the reduced distortion of the particle model with the ideal heterogeneous energy landscape with time (Fig. **4c**). The learning model was insensitive to the order of observations, recovering the same landscape shape and overall mean distortion regardless of observation order (Fig. **S3**).

Heterogeneous learning in neural networks

We next confirmed that similar trends emerged when considering synaptic heterogeneity in a full neural neural network model, which we have previously showed can be reduced asymptotically to the particle mdoels that preceded [16, 34]. Using a neural field model, in which excitatory neurons are tuned to preferential activation for particular feature values, we describe the neural activity at a particular location and time as

$$du(x,t) = [-u(x,t) + \int_{-\pi}^{\pi}(1 + A_n\cos(ny))w(x-y)f(u(y,t))dy]dt + \epsilon dW(x,t) + I(x,t)$$

where $u(x,t)$ is the average activity at location x and time t, $f(u)$ describes the firing rate nonlinearity, $\epsilon dW(x,t)$ describes spatiotemporal noise which is white in time and correlated in space, and I is a spatiotemporal input. Synaptic coupling $w(x-y)$ between locations x and y combines both excitatory-excitatory coupling as well as the disynaptic loop from excitatory to inhibitory and back, so the effective weight function exhibits lateral inhibition common to ring/bump attractor models [37]. Coupling between excitatory neurons W_{exc} (W_{ee}) is strongest to neurons with similar preferential tuning. Inhibitory-excitatory feedback W_{inh} ($W_{ei} * W_{ie}$) is wider than W_{exc} connections. The function $1 + A_n\cos(ny)$ introduces periodic spatial heterogeneity (learned or prescribed) that acts on the synaptic weighting, with A_n describing the amplitude and n defining the number of peaks. The spatial synaptic footprint emanating from each point y is the same when $A_n \equiv 0$ (homogeneous) and varies with $A_n > 0$ (heterogeneous, e.g., $n = 4$). The peak spatial synaptic footprint

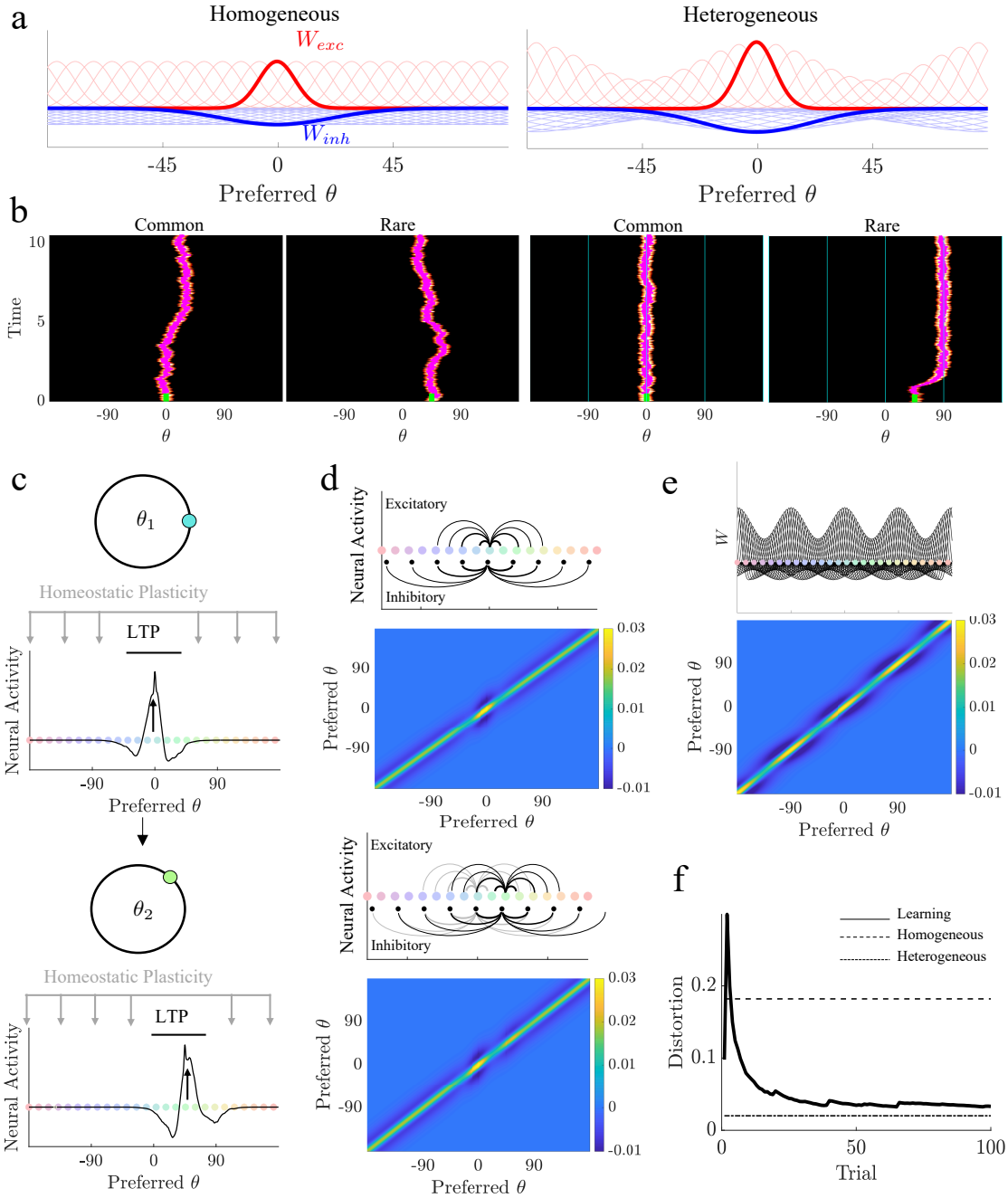

Figure 5: Experience-dependent modulation via long-term potentiation and homeostatic plasticity reduces distortion of encoded stimulus values. **a.** Localized excitatory (W_{exc}) and inhibitory (W_{inh}) synaptic weights associated with preferred item features in a neural field encoding a delayed estimate. Example of of synaptic weight originating from neuron with preference $\theta = 0$ shown in bold. Heterogeneous neural networks are modulated so synaptic footprints originating from peaks of $P_{env}(\theta)$ are stronger. Parameters used: $A_{inh} = 0.35, w_{inh} = 3, s_{het} = 0.4, n = 4$. **b.** Example bumps of sustained neural activity over 10s delay period originating at common and rare targets in a homogeneous (left) and heterogeneous (right) case. Cyan traces in heterogeneous plots denote attractor locations with enhanced synaptic weights. Stimulus input value shown in green. Parameters used: $A_{inh} = 0.35, W_{inh} = 3, s_{het} = 0.4, n = 4, T_{inp} = 0.5s, T_{delay} = 10s, h = 0.1, D = 0.5, A_{inp} = 1$. **c.** Experience-dependent learning results from long-term potentiation (LTP) of neurons with a preference for the previous target value and homeostatic reduction of connectivity elsewhere. **d.** Learning breaks the symmetry of the spatially-dependent weight kernel, creating enhanced peaks originating from neurons activated more across trials. **e.** After many trials, the weight matrix recovers the static heterogeneous synaptic structure. **f.** Average distortion over time decreases as the experience-dependent neural field model learns the environmental distribution. Both the final learning model's and heterogeneous model's mean distortion are significantly less than that of the homogeneous model (bootstrapping, $p < 0.05$). Parameters used: $A_{inh} = 0.3, W_{inh} = 3, s_{het} = 0.4, n = 4, T_{inp} = 0.5s, T_{delay} = 10s, h = 0.1, D = 0.5, A_{inp} = 1, s_{learn} = 0.1$, $N_{\text{boot}} = 1000$.

is chosen to be strongest at the peaks of the environmental distribution, resulting in locations (attractors) with stronger likelihood of activation (Fig. **5a**). The effects of heterogeneous synaptic weighting can be observed in the wandering patterns of a bump neural activation. In the homogeneous case, bumps wander freely are are equally likely to move any direction. Bumps in heterogeneous networks drift toward attractors (Fig. **5b**), reducing the likelihood they stray from common targets (near attractors) introducing larger drift and errors for rare targets.

We proposed a mechanistic learning rule by which experience modulates synaptic heterogeneity via potentiation of synapses emanating from neurons encoding prior stimulus values [16, 26], while homeostatic plasticity reduces synaptic weight strength elsewhere (Fig. **5c**). This modulation leads to an increase in connectivity strength at the prior target location (Fig. **5d**). Updates occur iteratively, so synaptic plasticity modulates weight functions across long timescales to reflect heterogeneities in the distribution of presented stimulus values (Fig. **5e**). As with the particle model, the neural network with learning maintains working memory representations whose distortion is reduced, compared to homogeneous networks, as the network learns (Fig. **5f**; bootstrapping, $p < 0.05$).

Subjects' behavior shows hallmarks of learning

We considered how our static and learning models compared to human behavior. Using subjects' delayed estimates from trials drawn from a color report task with a heterogeneous distribution of colors [15], we asked if subject responses were best fit to homogeneous, static heterogeneous, or learning dynamical models. Subjects were cued with two items (the target and distractor) and asked to respond with the color of one item after a short (0.5s) or long (4s) delay. Item colors were selected from either a heterogeneous (4-peak) input distribution or a uniform distribution. Heterogeneous color distribution peak positions were selected randomly for each subject (Fig. **6a**).

We compared responses to eight different particle models from three model classes. Our baseline model was a particle model with a homogeneous (flat) energy landscape. We also considered three static heterogeneous models: 1. a model whose energy landscape had a single frequency (static heterogeneous model), with free parameters for the amplitude, number of wells, and diffusion coefficient. The location of the wells were evenly spaced about the circle and oriented based on the subject's assigned offset; 2. a variation of the static heterogeneous model which allowed the energy wells to be offset from the environmental distribution peaks (Offset Heterogeneous model); and 3. a dual frequency model, in which the landscape incorporated two amplitudes and numbers of wells (Dual Heterogeneous model) (Fig. **6b**).

We also considered four learning models with two different forms: those that updated the energy landscape based only on the target (Target Only), and those that updated the potential landscape based on both items (Target + Distractor). In addition, we varied the prior to consider subjects' innate biases to the task colors. In one case, we assumed no inclusion of previous biases, beginning with a homogeneous landscape (Flat Prior). In the other form, we assumed subjects retained their innate color biases and initialized the model with a heterogeneous energy landscape centered at the peak locations of established population biases (Heterogeneous Prior) (Fig. **6c**).

To identify the model that best matched each subject's responses, we applied 5-fold cross-validation, computing the mean squared error between the responses and simulated data across possible parameter sets for each model. We found

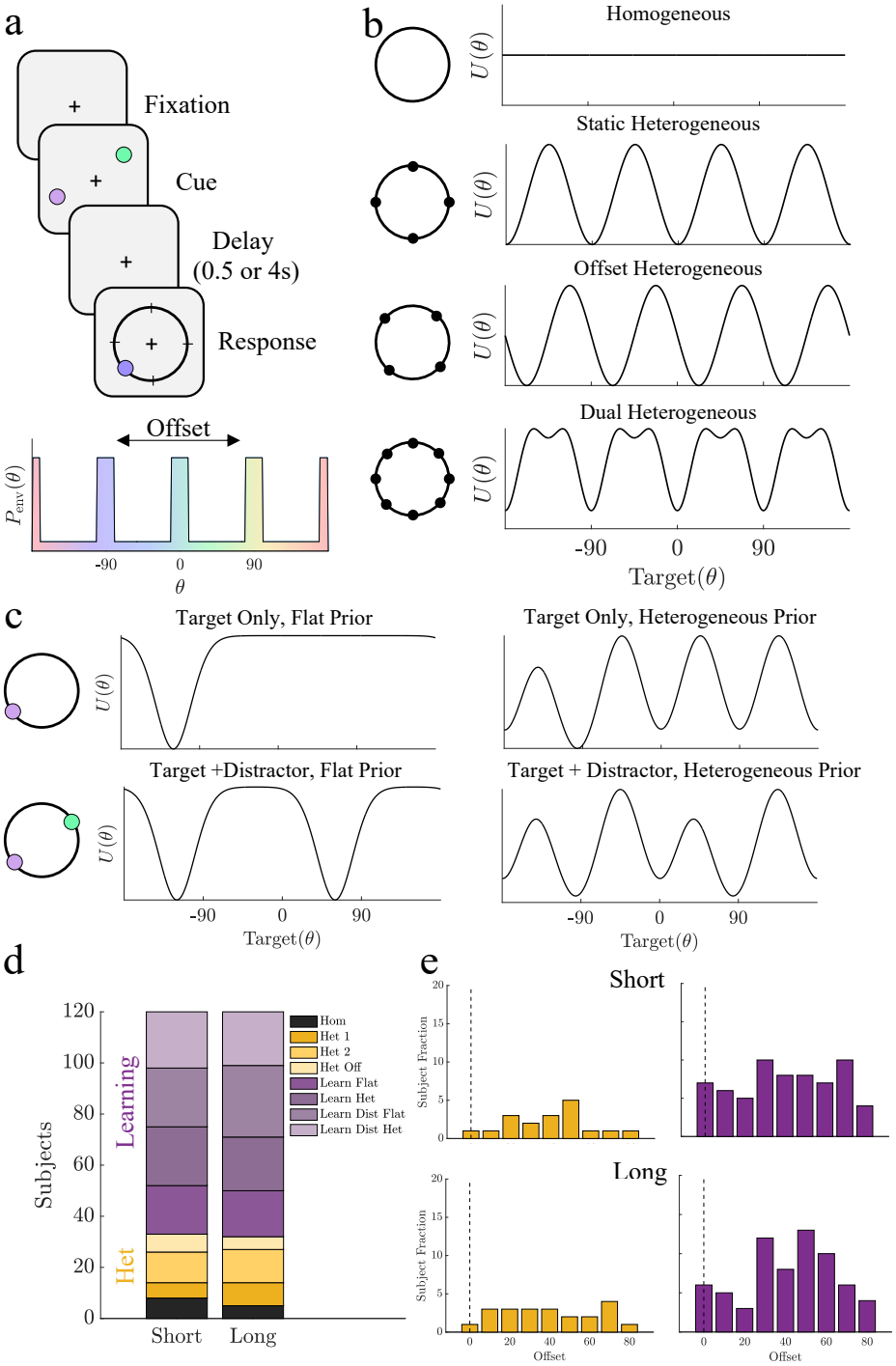

Figure 6: Subject responses to heterogeneous distributions are best replicated by static and dynamic heterogeneous models. **a.** Experiment 2 from [15] in which subjects were shown two items, each of which could be drawn from a heterogeneous distribution whose peaks were evenly distributed but randomly offset around the circle for each subject. Subjects were prompted to respond with the identifying color of one item. **b.** Static particle models used. Homogeneous included one free parameter, diffusion. Static Heterogeneous includes three free parameters: amplitude, number of wells, and diffusion. Offset Heterogeneous includes one additional parameter for offset. Dual heterogeneous considers five parameters: amplitude and number of wells for the first component, amplitude and number of wells for the second component, and diffusion. **c.** Learning particle models. Each updates iteratively based on three parameters: width of the bump, depth of the bump, and diffusion. Target-Only learning incorporated only the target prompted for response, and Target+ Distractor incorporates both items. Priors refer to initial landscape, beginning either with a homogeneous (flat) landscape or a heterogeneous landscape that matches the human population biases. **d.** Number of subject best matched to each model type for short and long delay periods. **e.** Assigned offsets for subjects best matched to static heterogeneous models (yellow) and learning models (purple). Dashed lines shows humans base bias location.

that a majority of the subjects (93% of subjects in short trials and 95% of subjects in long trials) were best described by learning and heterogeneous models, with most matching learning models (Fig. **6d**). We compared environmental (stimulus) peak offsets for each subject who was matched to to either the static heterogeneous or learning models. As compared to the original human bias, subjects best fit to learning models had assigned offsets that were centered away from the cardinal directions (away from their innate biases). Subjects best fit to static heterogeneous models showed no preference for particular offset values (Fig. **6e**). This finding suggests that many subjects confronted with a heterogeneous environment that differed from their bias applied experience-dependent learning to update their expectations about the environment as new information is acquired.

Discussion

Here, we have shown that systematic biases observed in human subject working memory can be attributed to environmental experience, especially with systematic variations in the frequency of stimulus feature values. Our work extends previous findings on working memory bias by proposing an experience-dependent learning mechanism as the basis for systematic errors that reflect heterogeneity in the environment. Beginning with a simplified model of delayed-estimate degradation, our particle model confirms that synaptic heterogeneity can stabilize memory and reduce errors made in estimating common feature values at the expense of larger errors for rare feature values. When attractor locations are matched to peaks in the feature value distribution, average memory distortion is reduced. Experience-dependent learning of the environment can be implemented neuromechanistically via long-term potentiation that enhances recurrent excitation for neurons encoding more common stimulus values with homeostatic plasticity regulating connectivity across the neural population. Finally, subjects confronted with heterogeneous environments displayed responses matched to parametric encoders with heterogeneous errors, with most subjects matched to learning models, particularly if task environments are not well matched to their prior beliefs. Thus, subjects confronted with environments that deviate from their priors appear to dynamically update their beliefs based on experience, supporting our hypothesis that systematic biases are learned by experience-dependent plasticity.

Our work supports previous findings that heterogeneity in working memory reduces variability, even when confronted with uniform environments [15, 25–27], and extends these findings to consider what is optimal for a non-uniform environment. While our models support the hypothesis that network heterogeneity is most beneficial when it mimics the environment, others have shown repulsive effects when implementing heterogeneous priors and efficient coding practices [5, 38]. Such "anti-Bayesian" biases can emerge from an alternative model in which environmental distributions are learned by redistributing the frequency of neural tuning rather than via synaptic plasticity. Having more neurons with stimulus preference close to the common feature values means the posterior estimate of stimuli is biased away from common stimulus value due to there being higher variance toward rare feature values [38]. However, the literature surrounding long-term biases in humans is mixed [22, 39–41], suggesting subjects do not necessarily take optimal approaches or necessarily the same approach to each situation. Our work supports this hypothesis by showing that subjects' responses in heterogeneous environments can be matched to a number of different models, many of which produce

suboptimal results.

For example, subjects' use of the distractor item as part of their updating procedure suggests that experience-dependent updates could occur during stimulus observation, rather than after their response, as has been suggested for short-term serial biases [7, 16]. Representations of memoranda in multiple item working memory tasks have also been shown to interact, sometimes causing additional errors in memory [19, 42, 43], but also reducing cardinal bias [44], suggesting that future work may consider how multi-item working memory tasks impact experience-dependent learning of heterogeneous environments. Notably, multi-items were presented sequentially in [44], implying that both short-term plasticity rules [16] and multi-item interactions, such as swapping errors [45], should be explored in the future.

While it is unsurprising that subject responses were best matched to heterogeneous models in the task from [15], we anticipated that static models would be more representative of subjects whose heterogeneous environments were more similar to their initial biases (offsets closer to the cardinal directions), requiring less updating to their environmental beliefs. In contrast, subjects who were best described by static models had offsets that appeared uniformly distributed, while a majority of the subjects described by learning models were exposed to environmental distributions with offsets that deviated from their biases. It is unclear whether subjects matched to static models were resistant to learning the new environment or were not provided sufficient experience (i.e., trials) to adapt, given that subjects experienced interspersed heterogeneous/homogeneous environments. Future studies could investigate the rate of learning when subjects are presented with a novel heterogeneous environment to identify the causes for subject-model variability.

This work establishes a mechanistic hypothesis for the impacts of environmental inference on working memory. Our model supports recent findings on training-induced changes in prefrontal cortex [46], and suggests that learning over longer timescales can have substantial stimulus-specific impacts in working memory.

Materials and Methods

Particle Model

All particle models with frozen energy landscapes had the same potential structure

$$U(\theta) = -\frac{A}{n}\cos(n\theta),$$

in which A describes the amplitude and n describes the number of wells (attractors). The homogeneous model is recovered when taking $A = 0$. Particle movement was simulated using a stochastic differential equation

$$d\theta(t) = -U'(\theta(t))dt + \sigma dW(t),$$

which incorporated noise as a Wiener process and in which the values for θ were discretized to 1 degree ($\pi/180$ radian) bins and time was discretized to 10ms bins. To recover the effects of drift alone $-U'(\theta(t))dt$, we removed the noise component by setting $\sigma = 0$.

Distortion

The mean distortion for the system is given by

$$\bar{d} = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} P(\theta_{env}) P(\theta_{resp}|\theta_{env})(1 - \cos(\theta_{resp} - \theta_{env})) d\theta_{env} d\theta_{resp},$$

which can be approximated by simulations sampled from the environmental distribution P_{env} ($Nsim = 1e5$).

The mean distortion for a particular value of θ, $D(\theta)$ was computed using simulations sampled at θ. To compute the stimulus-specific distortion for a given particle model and environment, θ was binned and simulations were used to compute $D(\theta)$ for the bin. Bootstrapping procedures were used to resample distortion and compute the standard deviations.

To compute the diffusion-only distortion for heterogeneous models, simulations were sampled at a single attractor/common stimulus value ($\theta = 0$), while drift-only dynamics were simulated with $\sigma = 0$.

Conditional and Marginal Distributions

The conditional probability $P(\theta_{resp}|\theta_{env})$ was computed by simulating the distribution of responses (particle end locations) for each discretized θ_{env} value. Marginal distributions of the response $P(\theta_{resp})$ were computed from the discrete conditional probability solutions.

Experience-dependent particle model

The dynamic particle model updates its well structure based on experienced stimuli using the equation

$$U_{est}^{n}(\theta) = U_{est}^{n-1}(\theta) + \frac{s(h - e^{\kappa \cos(\theta - \theta_{inp}) - 1})}{n}$$

which incorporates a von Mises distribution at the location of the stimuli θ with s as the scaling factor, h is the shift, and κ is the spread. The mean distortion and the particle landscape are updated iteratively on each trial, such that for each stimuli, the distortion is computed and included in the running average.

Neural Field Model

We use a neural field model based on the one from [47], in which the locations of neurons corresponds to their preferred stimulus.

$$du(x,t) = [-u(x,t) + \int_{-\pi}^{\pi} (1 + A_n \cos(ny)) w(x-y) f(u(y,t)) dy] dt + \epsilon dW(x,t) + I(x,t)$$

in which $u(x,t)$ is the mean activity at the location $x \in [-L, L]$ which has been discretized and scaled such that $x = x * L_{sc}$ to increase resolution. Connectivity across the grid is described by $w(x-y)$ which is described by

$$\exp(-(x-y)^2) - A_{inh}\exp(-(x-y)^2/w_{inh})$$

combining both local excitation and broad inhibition. Where A_{inh} is the strength of inhibition and w_{inh} describes the inhibitory spread.

Static heterogeneity is incorporated by $(1 + A_n(\cos(ny))$ and the firing rate nonlinearity $f(u(y,t))$ is taken to be a Heaviside function

$$H(u-\theta_{fr}) = \begin{cases} 1 & u > \theta_{fr} \\ 0 & u < \theta_{fr} \end{cases}$$

in which θ_{fr} describes the firing rate threshold.

Noise $\epsilon dW(x,t)$ is modeled as weak multiplicative noise driven by a spatially-dependent, white-in-time, Wiener process using the spatial filter

$$F = \sqrt{\epsilon}\exp(-|x-y|)\sqrt{(}dx),$$

and ϵ describes the noise strength.

Input to the system $I(x,t)$ is Gaussian, at location x_0,

$$I = I_0\exp(-(x-x_0)^2/2/I_s^2),$$

where I_0 is the strength of the input and I_s is the spread of the input.

Neural activity is updated iteratively and the bump's centroid is identified as the location with maximal activity. Parameters were selected such that less than 10% of all trials resulted in the bump extinguishing prior the end of the delay period. Responses for each trial were reported as the location of the center of mass at the end of the delay period.

See [16, 34] for details on asymptotically reducing the neural field model to an associated particle model with a heterogeneous energy landscape.

Plasticity rules in neural field model

To include experience-dependent learning into our neural field model, we replace the static heterogeneity with a learning function $(1+A(s(y,t))$ in which the function $s(y,t)$ updates with each trial based on neural activation while the network receives input (i.e., at the stimulus location) $(u(x, t_0\bar{:} t-1))$.

$$s(y,t_0;n) = s(y,t_0;n-1) + (u(x,t_0\bar{;} n-1))\kappa_s$$

where κ_s is a scaling factor for long-term plasticity. This function is then normalized to implement homeostatic plasticity rules.

Human data

Data was taken from [15], experiment 2, with permission. A short summary of the task given here. Details found in [15]. The task was administered to 120 consenting subjects with normal color vision in Amazon Mechanical Turk who performed and achieved minimal engagement. Task structure included 2 items presented simultaneously with a delay of 500 ms (100 short trials) or 4000 ms (100 long trials) prior to the response being cued. Each item had a 50% chance of being drawn from the biased distribution. The biased distribution included 4 peaks spanning 20 degrees, equally spaced about the circle. The peaks were shifted at random for each subject.

Subject model fitting

We fit subject responses to 8 particle models:

1. **The homogeneous potential** (1 free parameter) in which the particle dynamics are only influenced by diffusion

$$d\theta(t) = \sigma dW(t).$$

2. **Static Heterogeneous** (3 free parameters) in which the particle is subject to drift and diffusion parameterized by the potential parameters A (amplitude) and n (number of wells) and noise σ,

$$d\theta(t) = -A\sin(n\theta - \theta_{off})dt + \sigma dW(t),$$

where θ_{off} is the offset assigned to a subject by the experiment (not fit).

3. **Offset Heterogeneous** (4 free parameters) includes all of the above parameters but incorporates a free parameter for the offset value θ^s_{off}, such that a subject could use a model that does not match the true environment θ_{off}.

4. **Dual Heterogeneous** (5 free parameters) assumes that subject response are generated from a particle model using an energy landscape with two frequencies (n_1 and n_2) with amplitudes A_1 and A_1, assuming the offset to be at the assigned location. This the particle dynamics are described by

$$d\theta(t) = -A_1\sin(n_1\theta - \theta_{off})dt - A_2\sin(n_2\theta - \theta_{off})dt + \sigma dW(t).$$

5-6. **Target-only Learning** (3 free parameters) assumes the energy landscape is updated on each trial using an inverted von Mises distribution with free parameters for κ and s. Noise is parameterized by σ as before. Two forms of initial landscapes were tested,

$$U_0(\theta) = -A\cos(n\theta).$$

First, we considered a model with an initially homogeneous energy landscape in which $n = 0$. The second model ini-

tially assumed a heterogeneous landscape with $n = 4$ (matched to the number of peaks in the environmental distribution) with an offset matched to the population biases previously established.

7-8. **Target + Distractor Learning** (3 parameters) models were implemented equivalently to the "Target-only" model but were updated with both target and distractor stimuli on each trial.

Models were fit to each subject's set of responses using a 5-fold cross-validation procedure performed for short and long delay trials separately. For each subject-model, we tested 100 parameter sets on 80% of the trials, running 100 simulations with each set of parameters for each trial. The mean squared error (MSE) between each simulated response and subject responses were computed and the parameter set with the lowest MSE was selected for that subject-model pair. We then simulated responses for the final 20% of trials and computed the MSE for these trials. This process was performed 5 times, testing all trials. The testing-set MSEs were then averaged and the model and parameter set with lowest mean MSE was selected for each subject.

Acknowledgements

Funding: This work is funded by National Institutes of Health grant 1R01EB029847-01.

Author Contributions:

Conceptualization: TLE, ZPK

Methodology: TLE, ZPK

Investigation: TLE

Visualization: TLE

Supervision: ZPK

Writing-original draft: TLE

Writing-review and editing: TLE, ZPK

Competing Interests: The authors declare that they have no competing interests.

Data and Materials Availability: The data used to generate the figures and code developed for all proposed models has been made available on github.com/teissa/HeterogeneousWorkingMemory.

References

[1] Ma, W. J., Husain, M. & Bays, P. M. Changing concepts of working memory. *Nature neuroscience* **17**, 347–356 (2014).

[2] Luck, S. J. & Vogel, E. K. Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in cognitive sciences* **17**, 391–400 (2013).

[3] Fougnie, D., Suchow, J. W. & Alvarez, G. A. Variability in the quality of visual working memory. *Nature communications* **3**, 1–8 (2012).

[4] Bays, P. M. Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience* **34**, 3632–3645 (2014).

[5] Taylor, R. & Bays, P. M. Efficient coding in visual working memory accounts for stimulus-specific variations in recall. *Journal of Neuroscience* **38**, 7132–7142 (2018).

[6] Bliss, D. P., Sun, J. J. & D'Esposito, M. Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific reports* **7**, 1–13 (2017).

[7] Barbosa, J. *et al.* Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. *Nature neuroscience* **23**, 1016–1024 (2020).

[8] Ploner, C. J., Gaymard, B., Rivaud, S., Agid, Y. & Pierrot-Deseilligny, C. Temporal limits of spatial working memory in humans. *European Journal of Neuroscience* **10**, 794–797 (1998).

[9] Schneegans, S. & Bays, P. M. Drift in neural population activity causes working memory to deteriorate over time. *Journal of Neuroscience* **38**, 4859–4869 (2018).

[10] Compte, A., Brunel, N., Goldman-Rakic, P. S. & Wang, X.-J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral cortex* **10**, 910–923 (2000).

[11] Goldman-Rakic, P. S. Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).

[12] Wang, X.-J. Synaptic reverberation underlying mnemonic persistent activity. *Trends in neurosciences* **24**, 455–463 (2001).

[13] Wimmer, K., Nykamp, D. Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature neuroscience* **17**, 431–439 (2014).

[14] Constantinidis, C. & Klingberg, T. The neuroscience of working memory capacity and training. *Nature Reviews. Neuroscience* **17**, 438–449 (2016).

[15] Panichello, M. F., DePasquale, B., Pillow, J. W. & Buschman, T. J. Error-correcting dynamics in visual working memory. *Nature communications* **10**, 1–11 (2019).

[16] Kilpatrick, Z. P. Synaptic mechanisms of interference in working memory. *Scientific Reports* **8**, 7879 (2018).

[17] Barbosa, J. *et al.* Interplay between persistent activity and activity-silent dynamics in prefrontal cortex underlies serial biases in working memory. *Nature neuroscience* **23**, 1016–1024 (2020).

[18] Papadimitriou, C., Ferdoash, A. & Snyder, L. H. Ghosts in the machine: memory interference from the previous trial. *Journal of neurophysiology* **113**, 567–577 (2015).

[19] Almeida, R., Barbosa, J. & Compte, A. Neural circuit basis of visuo-spatial working memory precision: a computational and behavioral study. *Journal of neurophysiology* **114**, 1806–1818 (2015).

[20] Bae, G.-Y., Olkkonen, M., Allred, S. R. & Flombaum, J. I. Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General* **144**, 744–763 (2015).

[21] Scocchia, L., Cicchini, G. M. & Triesch, J. What's "up"? Working memory contents can bias orientation processing. *Vision Research* **78**, 46–55 (2013).

[22] Girshick, A. R., Landy, M. S. & Simoncelli, E. P. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience* **14**, 926–932 (2011).

[23] Heider, E. R. Universals in color naming and memory. *Journal of Experimental Psychology* **93**, 10–20 (1972).

[24] Hansen, B. C. & Essock, E. A. A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes. *Journal of Vision* **4** (2004).

[25] Renart, A., Song, P. & Wang, X.-J. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron* **38**, 473–485 (2003).

[26] Kilpatrick, Z. P., Ermentrout, B. & Doiron, B. Optimizing Working Memory with Heterogeneity of Recurrent Cortical Excitation. *The Journal of Neuroscience* **33**, 18999–19011 (2013).

[27] Pollock, E. & Jazayeri, M. Engineering recurrent neural networks from task-relevant manifolds and dynamics. *PLOS Computational Biology* **16** (2020).

[28] Louie, K. & Glimcher, P. W. Efficient coding and the neural representation of value. *Annals of the New York Academy of Sciences* **1251**, 13–32 (2012).

[29] Goldman-Rakic, P. S. Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates. *Progress in brain research* **85**, 325–336 (1991).

[30] Bays, P. M., Catalao, R. F. & Husain, M. The precision of visual working memory is set by allocation of a shared resource. *Journal of vision* **9**, 7–7 (2009).

[31] Burak, Y. & Fiete, I. R. Fundamental limits on persistent activity in networks of noisy neurons. *Proceedings of the National Academy of Sciences* **109**, 17645–17650 (2012).

[32] Itskov, V., Hansel, D. & Tsodyks, M. Short-term facilitation may stabilize parametric working memory trace. *Frontiers in computational neuroscience* **5**, 40 (2011).

[33] Schapiro, K., Josić, K., Kilpatrick, Z. P. & Gold, J. I. Strategy-dependent effects of working-memory limitations on human perceptual decision-making. *Elife* **11**, e73610 (2022).

[34] Kilpatrick, Z. P. & Ermentrout, B. Wandering bumps in stochastic neural fields. *SIAM Journal on Applied Dynamical Systems* **12**, 61–94 (2013).

[35] Lindner, B., Kostur, M. & Schimansky-Geier, L. Optimal diffusive transport in a tilted periodic potential. *Fluct Noise Lett* **1**, R25–R39 (2002).

[36] Koulakov, A. A., Raghavachari, S., Kepecs, A. & Lisman, J. E. Model for a robust neural integrator. *Nature neuroscience* **5**, 775–782 (2002).

[37] Amari, S.-i. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological cybernetics* **27**, 77–87 (1977).

[38] Wei, X.-X. & Stocker, A. A. A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature Neuroscience* **18**, 1509–1517 (2015).

[39] Bae, G.-Y., Olkkonen, M., Allred, S. R., Wilson, C. & Flombaum, J. I. Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision* **14**, 7 (2014).

[40] Pratte, M. S., Park, Y. E., Rademaker, R. L. & Tong, F. Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance* **43**, 6–17 (2017).

[41] de Gardelle, V., Kouider, S. & Sackur, J. An oblique illusion modulated by visibility: Non-monotonic sensory integration in orientation processing. *Journal of Vision* **10**, 6 (2010).

[42] Bae, G.-Y. & Luck, S. J. Interactions between visual working memory representations. *Attention, Perception, & Psychophysics* **79**, 2376–2395 (2017).

[43] Golomb, J. D. Divided spatial attention and feature-mixing errors. *Attention, Perception, & Psychophysics* **77**, 2562–2569 (2015).

[44] Bae, G.-Y. Breaking the cardinal rule: The impact of interitem interaction and attentional priority on the cardinal biases in orientation working memory. *Attention, Perception, & Psychophysics* (2021).

[45] Bays, P. M. Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports* **6**, 19203 (2016).

[46] Tang, H. *et al.* Prefrontal cortical plasticity during learning of cognitive tasks. *Nature Communications* **13**, 90 (2022).

[47] Krishnan, N., Poll, D. B. & Kilpatrick, Z. P. Synaptic efficacy shapes resource limitations in working memory. *Journal of Computational Neuroscience* **44**, 273–295 (2018).

Supplementary Materials

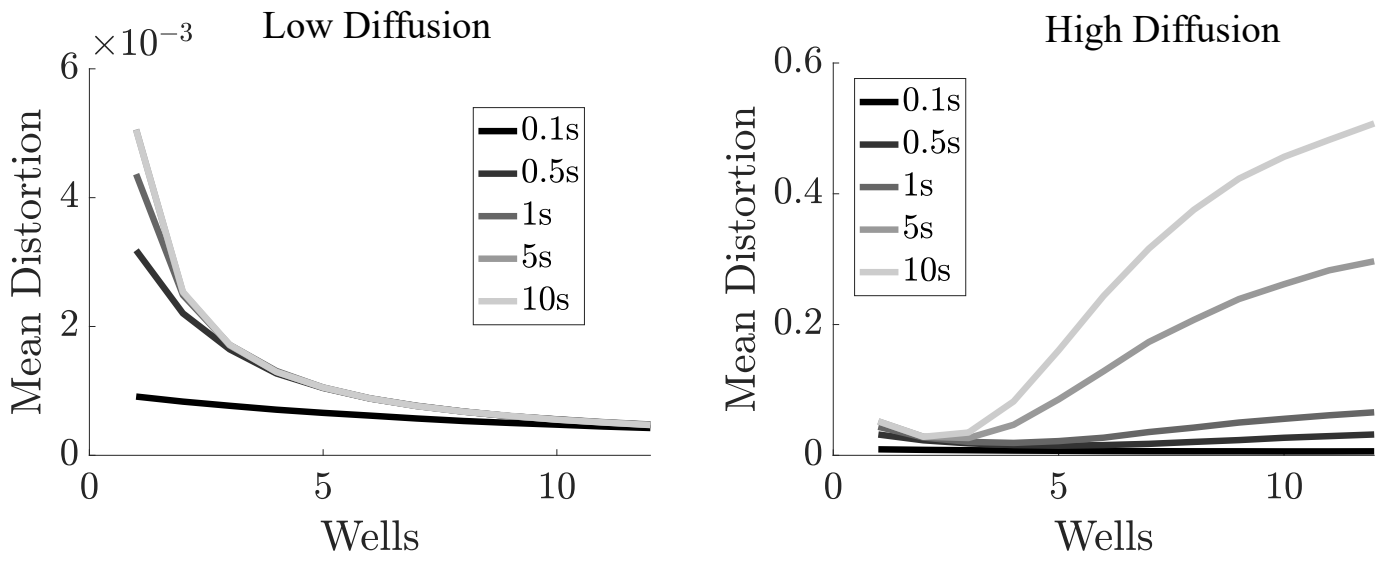

Figure S1: Mean distortion as amplitude and number of troughs was varied for four example delay periods. Optimal particle model identified based on minimum mean distortion (magenta dots). Top: Low diffusion ($D = 0.01$) leads to a optimal models with higher number of wells. Bottom: High diffusion ($D = 0.1$) leads to a optimal models with lower number of wells. Parameters used: $A = 1$, $N_{\text{Sims}} = 1e5$.

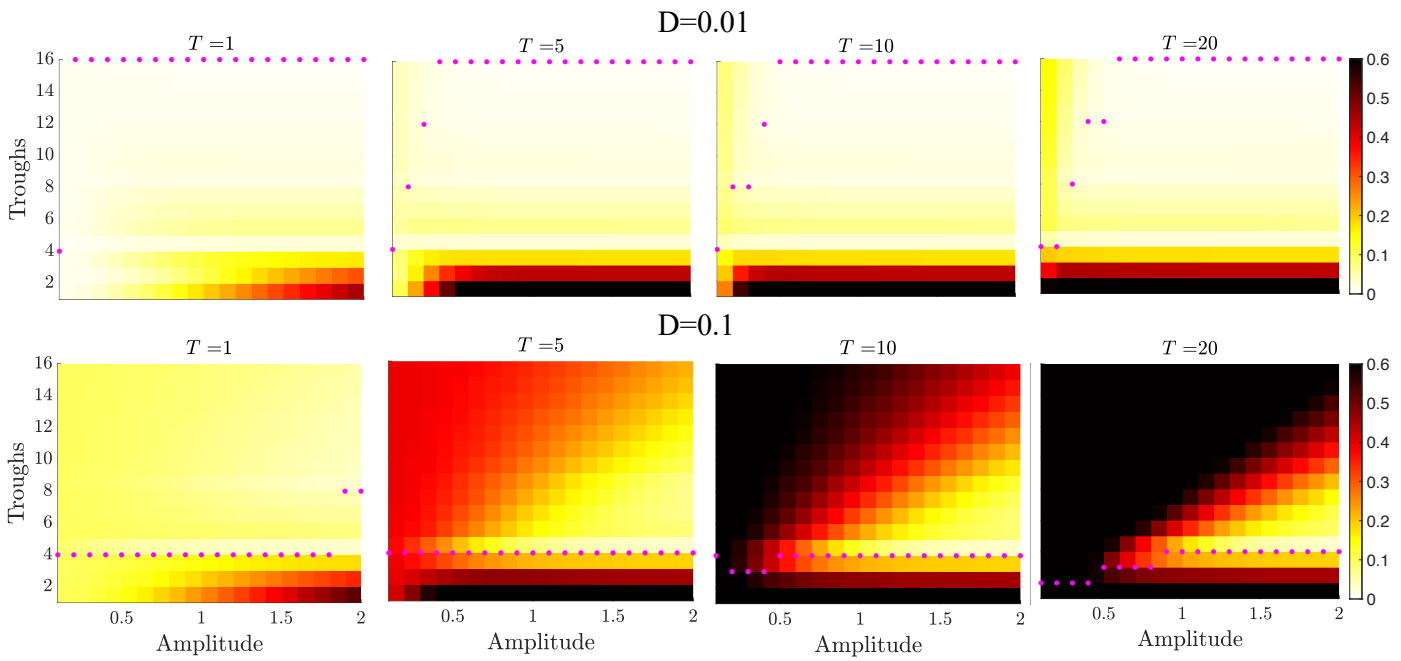

Figure S2: Mean distortion as amplitude and number of wells was varied for four example delay periods. Optimal particle model identified based on minimum mean distortion (magenta dots). Top: Low diffusion ($D = 0.01$) leads to a optimal models with higher number of wells. Bottom: High diffusion ($D = 0.1$) leads to a optimal models with lower number of wells. Parameters used: $A = 1$, $N_{\text{Sims}} = 1e5$.

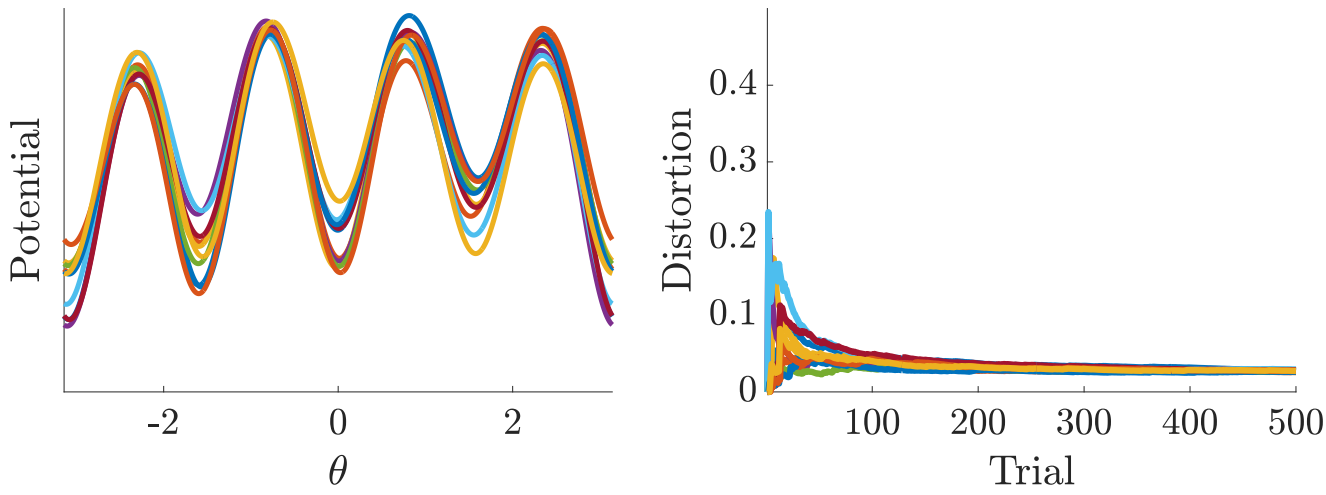

Figure S3: Ordering of observations in the learning particle model does not change the shape of the learned landscape or overall distortion. Left: 10 iterations of the learning model with the same observations but randomized permutations produce potential landscapes with the same shape but differing amplitudes. Right: 10 iterations of the learning model with no diffusion (drift only) and the same observations but randomized permutations show the same overall mean distortion after many trials with minor variations in the learning rate. Parameters used: $D = 0$, $T = 5$, $\kappa = 8$, $h = 0.25$, $s = 5$, $N_{\text{Trials}} = 500$.