From recency to the central tendency bias in working memory: a unifying attractor network model

Vezha Boboeva[1,2], Alberto Pezzotta[3,4], Athena Akrami[1,*,⋆], and Claudia Clopath[2,*,⋆]

[1]Sainsbury Wellcome Centre, University College London
[2]Department of Bioengineering, Imperial College London
[3]Gatsby Computational Neuroscience Unit, University College London
[4]The Francis Crick Institute
[*]These authors jointly supervised this work
[⋆]Corresponding authors

Monday 16th May, 2022

Abstract

The central tendency bias, or contraction bias is a phenomenon where the judgment of the magnitude of items held in working memory is biased towards the average of past observations. This phenomenon has been first described more than a century ago [1] and since then, has been replicated in various decision making tasks in humans [2–10], and rodents [11, 12]. Contraction bias is assumed to be an optimal strategy by the brain, given the noisy nature of working memory. From a Bayesian perspective [7], the progressive shift of the noisy memory towards the mean of a prior distribution built from past sensory experience helps with more accurate estimates of the memory. In this work, we propose an alternative account, via short-term history biases (serial dependence) [12–15]. Our model is motivated and inspired by recent results from an auditory delayed-discrimination task in rats, where the posterior parietal cortex (PPC) has been shown to be critical to these memory effects [12]. The dynamics of our model suggests that contraction bias can emerge as a result of a volatile working memory content which makes it susceptible to shifting to the past sensory experience. The errors, at the level of individual trials, are sampled from the full distribution of the stimuli, and are not due to a gradual shift of the memory towards the distribution's mean. Our model explains both short-term history biases, as well as contraction bias towards the sensory mean for the averaged performance. The results are consistent with the role of the PPC in encoding such sensory history biases, and provide predictions of performance across different stimulus distributions and timings, delay intervals, as well as neuronal dynamics in putative working memory areas.

Introduction

Working memory is the ability to hold and manipulate information for a short period of time. This ability subserves different cognitive and executive behaviors and has been studied in a variety of paradigms, including parametric working memory (PWM) tasks [16]. Such tasks involve the sequential comparison of two graded stimuli that differ along a physical dimension and are separated by a delay interval of a few seconds (Fig. 1 A). This paradigm has been studied in rodents, non-human primates, as well as humans, in a variety of modalities, ranging from tactile to auditory [16, 12, 7], and has offered empirical insights into the mechanisms involved in sensory working memory. One key feature, emerging from a multitude of PWM studies, is the effect known as "contraction bias", where the averaged performance is as if the memory of the first stimulus progressively shifts towards the center of a prior distribution built from past sensory history (Fig. 1 B) [1, 6, 3, 4, 13, 14]. Contraction bias, however, is not the only type of bias emerging in delayed comparison tasks. Biases towards the most recent sensory stimuli (previous immediate trials), similar to serial dependence have been documented [12, 13, 17–23, 14, 24].

Recently, it has been shown that optogenetically inactivating a region of rats' posterior parietal cortex (PPC) leads to a change in their performance in an auditory delayed comparison task, compatible with the elimination of sensory history biases [12]. Interestingly, other behavioral components, including working memory of immediate sensory stimuli (in the current trial), remain intact. This suggests that the sensory history information, provided by the PPC, may be relayed to another downstream region where it is then used and integrated with the working memory of the current stimulus. In the same study, additionally, it has been shown that a sub-population of PPC neurons carries more information about the stimuli presented during previous trials relative to the current trial, suggesting that the PPC can process information over a slower timescale than is necessary for solving the task. In a recent study, it was shown that contrary to choice biases, contraction bias cannot be modified using feedback protocols. This suggests the existence of unsupervised network mechanisms that give rise to contraction bias in humans [25].

Building on these findings, we propose a two-module model aimed at describing contraction bias as well as the short-term sensory history effects observed in the delayed comparison tasks [12]. Our model is comprised of a putative working memory (WM) network as well as a separate network modeling the PPC. Each network represents a continuous (bump) attractor. Given the finding that PPC neurons carry more information about stimuli presented during previous trials, our PPC module integrates inputs over a slower timescale relative to the WM network, and incorporates firing rate adaptation. Moreover, we assume that any information needed to solve the comparison task is read out from the WM network, and not the PPC network. We find that both contraction bias and short-term sensory history effects emerge in the WM network as a result of inputs from the PPC network. Importantly, we find that these effects do not occur due to separate mechanisms. Rather, contraction bias emerges as a statistical effect of errors in working memory: when the first stimulus in memory is replaced by a random representation taken from the same distribution, the pattern of errors is consistent with contraction bias. In our model, these errors occur due to the persistence of the memory of stimuli shown in the preceding trials in the PPC, the integration of which disrupts the memory of stimuli in the current trial in the WM network. This gives rise to short-term history effects. As a result, our model makes different predictions on the shape the performance takes, depending on the distribution of the stimuli. Taken together, we show that the slower timescale of the PPC network, as well as adaptation, is sufficient to produce short-term memory effects as well as contraction bias in the performance of subjects in a parametric working memory task, and that these different biases can occur due to the same mechanism.

1 Results

1.1 The PPC as a slower integrator network

In order to study the biases in WM, we use two identical one-dimensional continuous attractor networks to model WM and PPC modules. Neurons are arranged according to their preferential firing locations in a continuous stimulus space, representing the amplitude of auditory stimuli. Excitatory recurrent connections between neurons are symmetric and a monotonically decreasing function of the distance between the preferential firing fields of neurons, allowing neurons to mutually excite one another; inhibition, instead, is uniform. Together, both allow a localized bump of activity to form and be sustained (Fig. 1 C and D) [26, 27]. Both networks have free boundary conditions. Neurons in the WM network receive inputs from neurons in the PPC coding for the same stimulus amplitude (Fig. 1 C). Moreover, building on experimental findings [28, 12], we designed the PPC network such that it integrates activity over a slower timescale compared to the WM network (Sect. 3.1). Moreover, neurons in the PPC are equipped with neural adaptation.

To simulate the parametric WM task, at the beginning of each trial, the network is provided with a stimulus s_1 for a short time via an external current I_{ext} as input to a set of neurons (see Tab. 1). Following s_1, after a delay interval, a second stimulus s_2 is presented (Fig. 1 D). The pair (s_1, s_2) is drawn from the stimulus set shown in Fig. 1 D, where they are all equally distant from the diagonal $s_1 = s_2$, and are therefore of equal nominal discrimination, or difficulty. The stimuli (s_1, s_2) are co-varied in each trial so that the task cannot be solved by relying on only one of the stimuli [29]. As in the study in Ref. [12] using an interleaved design, across consecutive trials, the inter-stimulus delay intervals are randomized and sampled uniformly between 2, 6 and 10 seconds.

We additionally include psychometric pairs (indicated in the box in Fig. 1 A) where the distance to the diagonal, hence the discrimination difficulty, is varied. The task is a binary comparison task that aims at classifying whether $s_1 > s_2$ or vice-versa. In order to solve the task, we record the activity of the WM network at two time points: slightly before and after the onset of s_2 (Fig. 1 D). We repeat this procedure across many different trials, and use the recorded activity to assess performance (see (Sect. 3.2) for details). Importantly, at the end of each trial, the activity of both networks is not re-initialized, and the state of the network at the end of the trial serves as the initial network configuration for the next trial (Fig. 1 D).

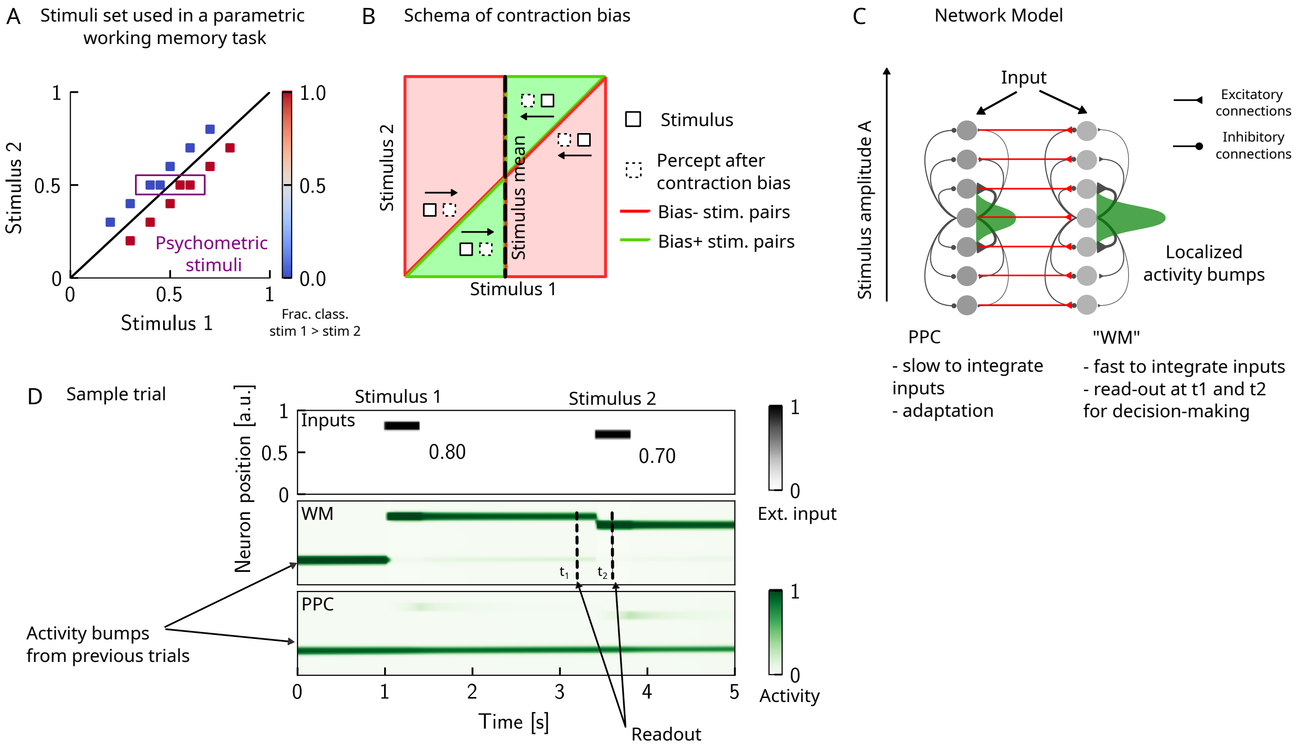

Figure 1: The PPC as a slower integrator network. (A) The stimulus set. In any given trial, a pair of stimuli corresponding to a square in this space is presented to the network with a delay interval separating them. The task is to compare the two stimuli, and report which one is larger. The stimuli are linearly separable, and stimulus pairs are equally distant from the $s_1 = s_2$ diagonal. Optimal performance corresponds to network dynamics from which it is possible to classify all the stimuli below the diagonal as $s_1 > s_2$ (shown in red) and all stimuli above the diagonal as $s_1 < s_2$ (shown in blue). An example of a correct trial can be seen in (D). In order to assay the psychometric threshold, several additional pairs of stimuli are included (purple box), where the distance to the diagonal $s_1 = s_2$ is systematically changed. The colorbar corresponds to the fraction classified as $s_1 > s_2$. **(B)** Schematics of contraction bias in delayed comparison tasks. Performance is a function of the difference between the two stimuli, and is impacted by contraction bias, where the base stimulus s_1 is perceived as closer to the mean stimulus. This leads to a better/worse (green/red area) performance, depending on whether this "attraction" increases (Bias+) or decreases (Bias-) the discrimination between the base stimulus s_1 and the comparison stimulus s_2. **(C)** Our model is composed of two modules, representing working memory (WM), and sensory history (PPC). Each module is a continuous one-dimensional attractor network. Both networks are identical except for the timescales over which they integrate external inputs; PPC has a significantly slower integration timescale and its neurons are additionally equipped with neuronal adaptation. The neurons in the WM network receive input from those in the PPC, through connections (red lines) between neurons coding for the same stimulus. Neurons (gray dots) are arranged according to their preferential firing locations. The excitatory recurrent connections between neurons in each network are a symmetric, decreasing function of their preferential firing locations, whereas the inhibitory connections are uniform (black lines). For simplicity, connections are shown for a single pre-synaptic neuron (where there is a bump in green). When a sufficient amount of input is given to a network, a bump of activity is formed, and subsequently sustained in the network when the external input is removed. This activity in the WM network is read out at two time points: slightly before and after the onset of the second stimulus, and is used to solve the comparison task. **(D)** The task involves the comparison of two sequentially presented stimuli, interleaved by a delay interval (top panel, black lines). The WM network integrates and responds to inputs quickly (middle panel), while the PPC network integrates inputs relatively slower (bottom panel). As a result, external inputs are enough to displace the bump of activity to the location of a new stimulus in the WM network, whereas the same does not always hold true for the PPC, as shown in this particular sample trial.

1.2 Contraction bias and short-term stimulus history effects as a result of PPC network activity

Probing the WM network performance on psychometric stimuli (Fig. 1 A, purple box, 10% of all trials) shows that the comparison behavior is not error-free, and that the psychometric curves (different colors) differs from the optimal step function (Fig. 2 A, green dashed line). Additionally, performance worsens as a function of the inter-stimulus delay interval [30]. The errors are caused by the displacement of the activity bump in the WM network, due to the inputs from the PPC network. These displacements in the WM activity bump can result in different outcomes: by displacing it *away* from the second stimulus, they either do not affect the performance or improve it (Fig. 2 B bottom panel, “Bias+”), if noise is present. Conversely, the performance can suffer, if the displacement of the activity bump is *towards* the second stimulus (Fig. 2 B top panel, “Bias-”). Performance on stimulus pairs that are equally distant from the $s_1 = s_2$ diagonal can be similarly impacted and the network produces a pattern of errors that is consistent with contraction bias: performance is at its minimum for stimulus pairs in which s_1 is either largest or smallest, and at its maximum for stimulus pairs in which s_2 is largest or smallest (Fig. 2 C) [7, 11, 12, 31].

Can the same circuit also give rise to short-term sensory history biases [12, 25]? We analyzed the comparison performance in the current trial conditioned on different stimulus pairs presented in the previous trial. As evident in Fig. 2 D, the modulation of performance by stimulus pairs presented in the previous trial reveals short-term sensory history biases, similar to what has been found experimentally [12]. We further quantified these history effects as well as how many trials back they extend to. We computed the bias by plotting, for each particular pair presented at the current trial, the performance as a function of the pair presented in the previous trial minus the mean performance over all previous pairs (Fig. 2 E) [12]. Independent of the current trial, the previous trial exerts an “attractive” effect, expressed by the positive slope of the line in Fig. 2 E: when the previous pair of stimuli is large, s_1 in the current trial is, on average, misclassified as larger than it is, giving rise to the attractive bias in the comparison performance; the converse holds true when the previous pair of stimuli happens to be small (Fig. 2 E, left panel). These effects extend to two trials back (Fig. 2 E, middle and right panels).

It has been shown that inactivating the PPC, in rats performing a delayed comparison task, markedly reduces the magnitude of the contraction bias, without impacting the non-sensory biases [12]. We assay the causal role of the PPC in generating the sensory history effects as well as contraction bias by weakening the connections from the PPC to the WM network, mimicking the inactivation of the PPC. In this case, we see that the performance for both the psychometric (orange curve, Fig. 2 F) as well as other pairs of stimuli is greatly improved (Fig. S3 A). The breakdown of the performance of the current trial conditioned on the specific stimulus pair preceding it reveals that the previous trial no longer exerts a notable modulating effect on the current trial (Fig. S3 B). Quantifying this bias by subtracting the mean performance over all of the previous pairs reveals that the attractive bias is virtually eliminated (orange curve, Fig. 2 G).

Together, our results suggest a possible circuit through which both contraction bias and short-term history effects in a parametric working memory task may arise. The main features of our model are two continuous attractor networks, both integrating the same external inputs, but operating over different timescales. Crucially, the slower one, a model of the PPC, includes neuronal adaptation, and provides input to the faster one, intended as a WM circuitry. In the next section, we show how the slow integration and firing rate adaptation in the PPC network give rise to the observed effects of sensory history.

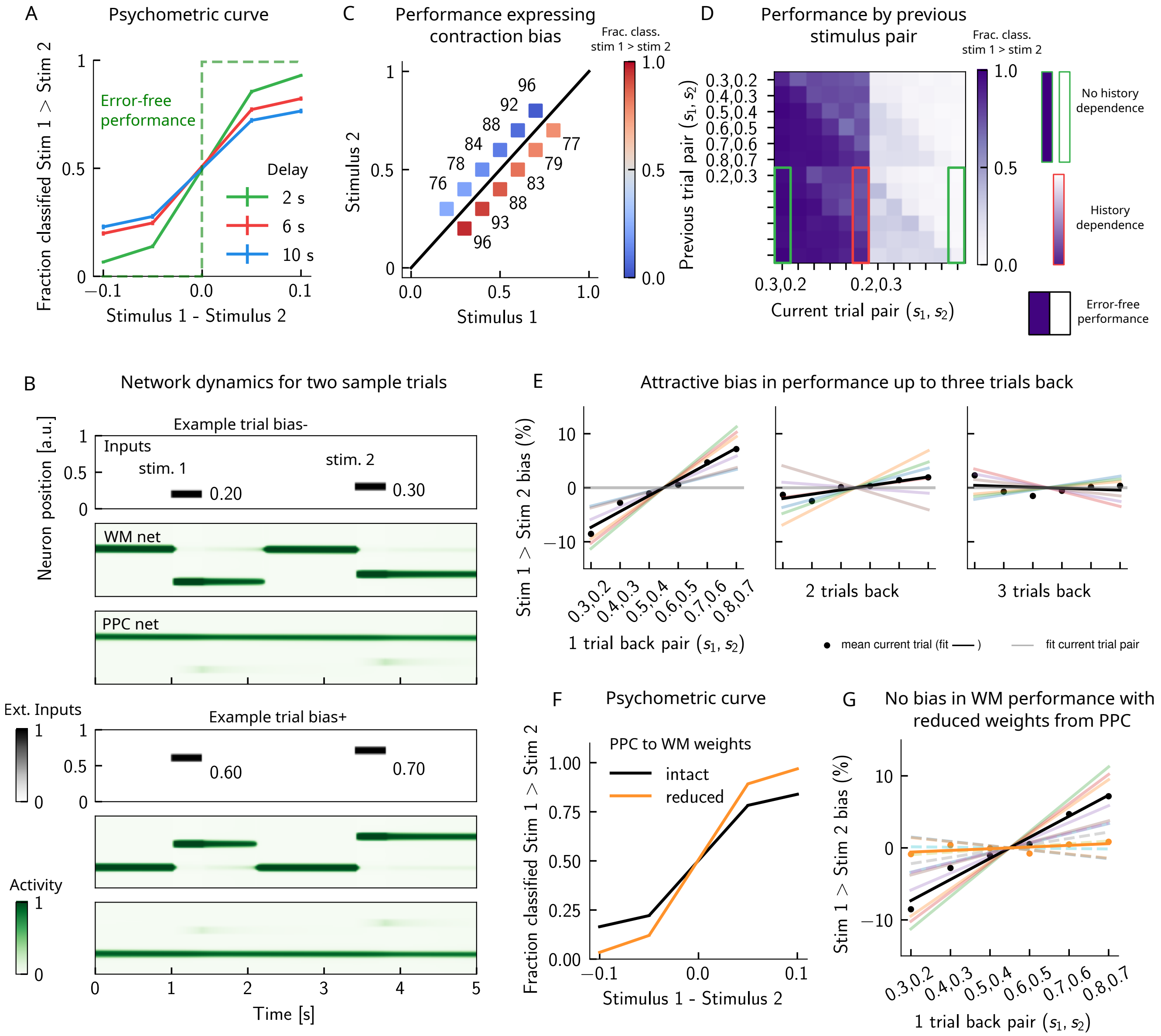

Figure 2: Contraction bias and short-term sensory history effects as a result of PPC network activity. (A) Performance of the network model for the psychometric stimuli (colored lines) is not error-free (green dashed lines). A shorter inter-stimulus delay interval is associated with a better performance. Errorbars (too small to note) correspond to the s.e.m. over different simulations. **(B)** Errors occur due to the displacement of the bump representing the first stimulus s_1 in the WM network. Depending on the direction of this displacement with respect to s_2, this can give rise to trials in which the comparison task becomes harder (easier), leading to negative (positive) biases (top and bottom panels). The top sub-panel displays the stimuli given to both networks in time, and the middle and bottom sub-panels show the activity of the WM and PPC networks (in green). **(C)** Comparison errors occur for equally discriminable stimuli ($|s_1 - s_2| =$ fixed). Performance is affected by contraction bias – a gradual accumulation of errors for stimuli below (above) the diagonal upon increasing (decreasing) s_1. Colorbar indicates the fraction of trials classified as $s_1 > s_2$. **(D)** Performance is affected by the previous stimulus pairs (modulation along the y-axis). The colorbar corresponds to the fraction classified $s_1 > s_2$. For readability, only some tick-labels are shown. **(E)** Left panel: bias, quantifying the (attractive) effect of the previous stimulus pairs. Colored lines correspond to linear fits of this bias for each pair of stimuli in the current trial. Black dots correspond to the average over all the current stimuli, and the black line is a linear fit. These history effects are attractive: the larger the previous stimulus, the higher the probability of classifying the first stimulus of the current trial s_1 as large, and vice-versa. Middle/right panels: same as the left panel, for stimuli extending two and three trials back. For this set of parameters (see Tab. 1), attractive effects extend to two trials back. **(F)** The performance of the network when the strength of the inputs from the PPC to the WM network are weakened (modelling the optogenetic inactivation of the PPC), is improved for psychometric stimuli (orange line), relative to the intact network (black line). **(G)** The attractive bias due to the effect of the previous trial stimulus pairs is eliminated, consistent with experimental results [12]. Dashed colored lines correspond to linear fits of this bias for each pair of stimuli in the current trial, the orange line is the mean over all the current trials.

1.3 Multiple timescales at the core of short-term sensory history effects

The activity bumps in the PPC and WM networks undergo different dynamics, due to the different timescales with which they integrate inputs, the presence of adaptation in the PPC, and the integration of incoming inputs from the PPC to the WM network. The WM network integrates inputs over a shorter timescale, and therefore the activity bump follows the external input with high fidelity (Fig. 3 A (purple bumps) and B (purple line)). The PPC network, instead, has a slower integration timescale, and therefore fails to sufficiently integrate the input to induce a displacement of the bump to the location of a new stimulus, at each single trial. This is mainly due to the recurrent inputs sustaining the bump being larger than the external stimuli that are integrated. The external input, as well as the presence of adaptation (Fig. S1 B and C) induce a small continuous drift of the activity bump that is already present from the previous trials (Fig. 3 A (pink bumps) and B (pink line)) [32]. The build-up of adaptation in the PPC network, combined with the global inhibition from other neurons receiving external inputs, can extinguish the bump in that location (see also Fig. S1 for more details). Following this, the PPC network can make a transition to an incoming stimulus position (that may be either s_1 or s_2), and a new bump is formed. The resulting dynamics in the PPC is a mixture of slow drift over a few trials, followed by occasional jumps (Fig. 3 A).

As a result of this dynamics, relative to the WM network, the activity bump in the PPC represents the stimuli corresponding to the current trial in a fewer fraction of the trials, and represents stimuli presented in the previous trial (Fig. 3 C). This yields short-term sensory history effects in our model (Fig. 2 D, and E), as input from the PPC lead to the displacement of the WM bump to other locations (Fig. 3 D). Given that neurons in the WM network integrate this input, once it has built up sufficiently, it can surpass the self-sustaining inputs from the recurrent connections in the WM network. The WM bump, then, can move to a new location, given by the position of the bump in the PPC (Fig. 3 D). As the input from the PPC builds up gradually, the probability of bump displacement in WM increases over time. This in return leads to an increased probability of contraction bias (Fig. 3 E), and short-term history (one-trial back) biases (Fig. 3 F), as the inter-stimulus delay interval increases.

Additionally, a non-adapted input from the PPC has a larger likelihood of displacing the WM bump. This is highest immediately following the formation of a new bump in the PPC, or in other words, following a "bump jump" (Fig. 3 F). As a result, one can reason that those trials immediately following a jump in the PPC are the ones that should yield the maximal bias towards stimuli presented in the previous trial. We therefore separated trials according to whether or not a jump has occurred in the PPC in the preceding trial (we define a jump to have occurred if the bump location across two consecutive trials in the PPC is displaced by an amount larger than the typical width of the bump (Sect. 3.1)). In line with this reasoning, only the set that included trials with jumps in the preceding trial yields a one-trial back bias (Fig. 3 G).

Removing neuronal adaptation entirely from the PPC network further corroborates this result. In this case, the network dynamics shows a very different behavior: the activity bump in the PPC undergoes a smooth drift (Fig. S2 A) [33]. In this regime, there are no jumps in the PPC, and the activity bump corresponds to the stimuli presented in the previous trial in a fewer fraction of the trials (Fig. S2 B), relative to when adaptation is present (Fig. 3 B). As a result, no short-term history effects can be observed (Fig. S2 C), even though a strong contraction bias persists (Fig. S2 D).

As in the study in Ref.[12], we can further study the impact of the PPC on the dynamics of the WM network by weakening the weights from the PPC to the WM network, mimicking the inactivation of PPC (Fig. 2 F and G, Fig. S3 A and B). Under this manipulation, the trajectory of the activity bump in the WM network immediately before the onset of the second stimulus s_2 closely follows the external input, consistent with an enhanced WM function (Fig. S3 C and D).

The drift-jump dynamics in our model of the PPC gives rise to short-term (notably one and two-trial back) sensory history effects in the performance of the WM network. In addition, we observe an equally salient contraction bias (bias towards the sensory mean) in the WM network's performance, increasing with the delay period (Fig. 3 E). However, we find that the activity bump in both the WM and the PPC network corresponds to the mean over all stimuli in only a small fraction of trials, expected by chance (Fig. 3 B, see Sect. 4.1 for how it is calculated). Rather, the bump is located more often at the current trial stimulus (s_1^t), and to a lesser extent, at the location of stimuli presented at the previous trial (s_2^{t-1}). As a result, contraction bias in our model cannot be attributed to the representation of the running sensory average in the PPC. In the next section, we show how contraction bias arises as an averaged effect, when single trial errors occur due to short-term sensory history biases.

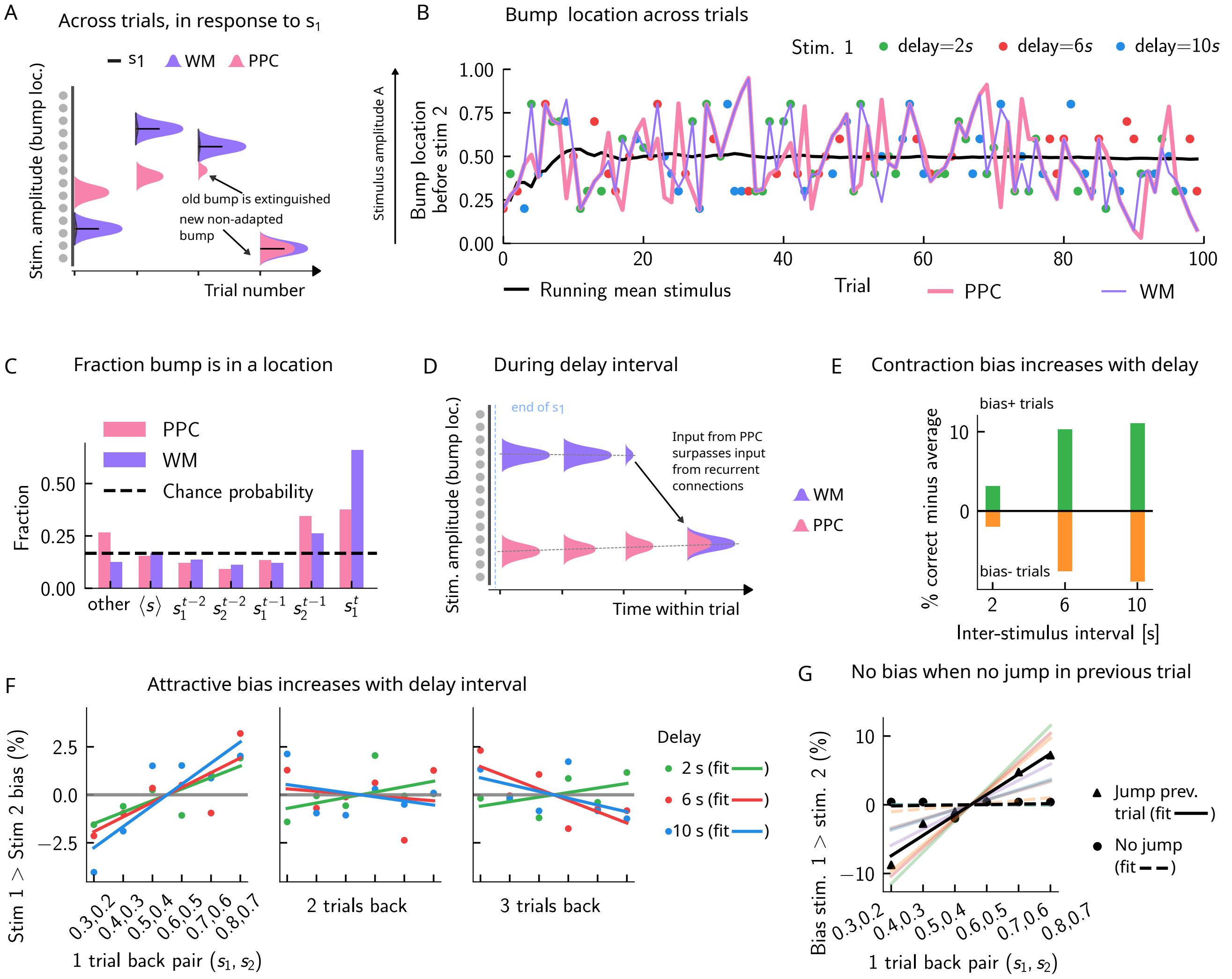

Figure 3: Multiple timescales at the core of short-term sensory history effects. (A) Schematics of activity bump dynamics in the WM vs PPC network. Whereas the WM responds quickly to external inputs, the bump in the PPC drifts slowly and adapts, until it is extinguished and a new bump forms. **(B)** The location of the activity bump in both the PPC (pink line) and the WM (purple line) networks, immediately before the onset of the second stimulus s_2 of each trial. This location corresponds to the amplitude of the stimulus being encoded, with lower amplitudes encoded towards the bottom and higher amplitudes towards the top. The bump in the WM network closely represents the stimulus s_1 (shown in colored dots, each color corresponding to a different delay interval). The PPC network, instead, being slower to integrate inputs, displays a continuous drift of the activity bump across a few trials, before it jumps to a new stimulus location, due to the combined effect of inhibition from incoming inputs and adaptation that extinguishes previous activity. **(C)** Fraction of trials in which the bump location corresponds to the base stimulus that has been presented (s_1^t) in the current trial, as well as the two preceding trials (s_2^{t-1} to s_1^{t-2}). In the WM network, in the majority of trials, the bump coincides with the first stimulus of the current trial s_1^t. In a smaller fraction of the trials, it corresponds to the previous stimulus s_2^{t-1}, due to the input from the PPC. In the PPC network instead, a smaller fraction of trials consist of the activity bump coinciding with the current stimulus s_1^t. Relative to the WM network, the bump is more likely to coincide with the previous trial's comparison stimulus (s_2^{t-1}). **(D)** During the inter-stimulus delay interval, in the absence of external sensory inputs, the activity bump in the WM network is mainly sustained endogenously by the recurrent inputs. It may, however, be destabilized by the continual integration of inputs from the PPC. **(E)** As a result, with an increasing delay interval, given that more errors are made, contraction bias increases. Green (orange) bars correspond to the performance in Bias+ (Bias-) regions, relative to the mean performance over all pairs (Fig. 1 A). **(F)** Longer delay intervals allow for a longer integration times which in turn lead to a larger frequency of WM disruptions due to previous trials, leading to a larger previous-trial biases ($2s$ vs. $6s$ vs. $10s$). Colored dots correspond to the bias computed for different values of the inter-stimulus delay interval, while colored lines correspond to their linear fits. **(G)** When neuronal adaptation is at its lowest in the PPC i.e. following a bump jump, the WM bump is maximally susceptible to inputs from the PPC. The attractive bias (towards previous stimuli) is present in trials in which a jump occurred in the previous trial (black triangles, with black line a linear fit). Such biases are absent in trials where no jumps occur in the previous trial (black dots, with dashed line a linear fit). Colored lines correspond to bias for specific pairs of stimuli in the current trial, regular lines for the jump condition, and dashed for the no jump condition.

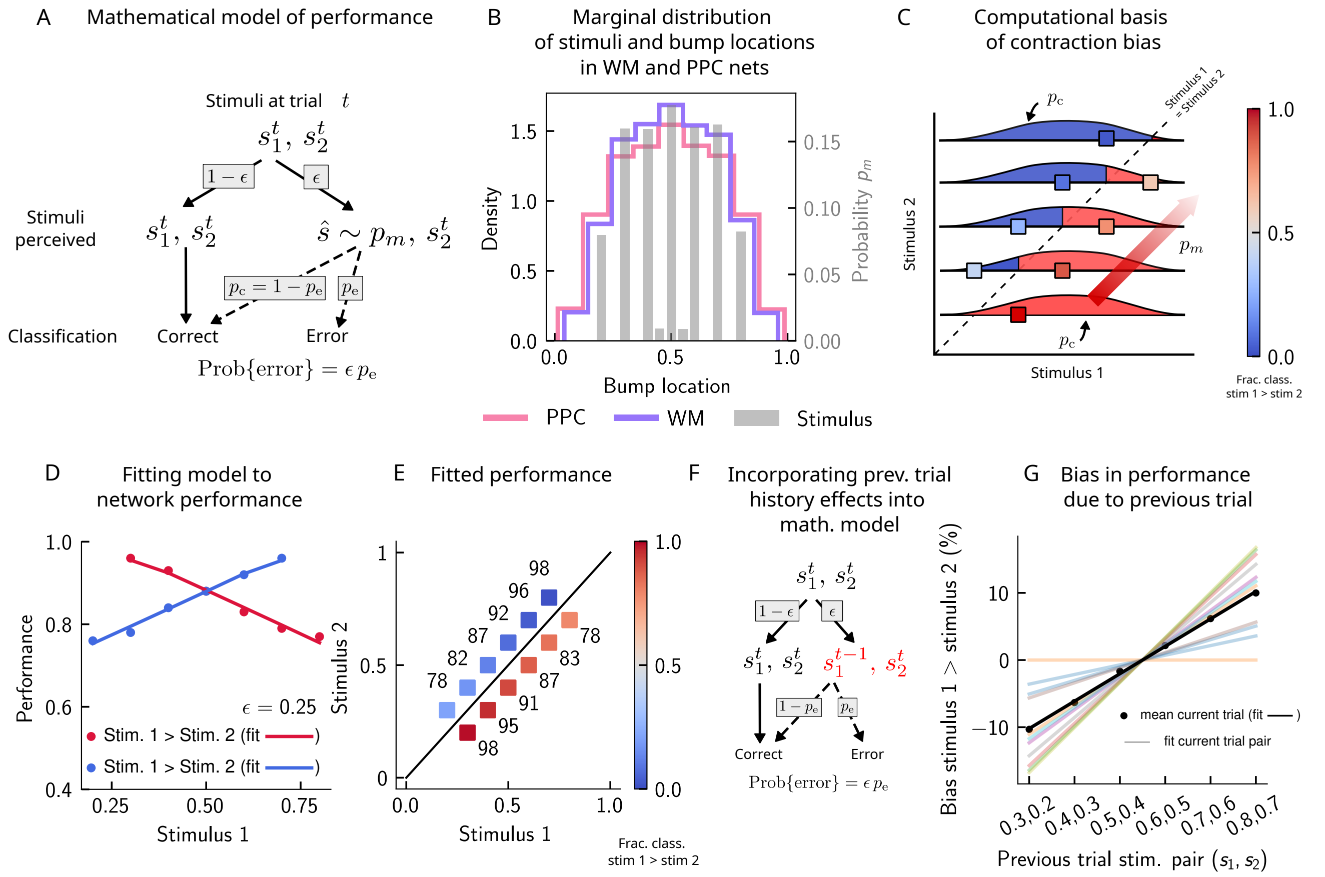

Figure 4: Errors are drawn from the marginal distribution of stimuli, giving rise to contraction bias. (A) A simple mathematical model illustrates how contraction bias emerges as a result of a volatile working memory for s_1. A given trial consists of two stimuli s_1^t and s_2^t. We assume that the encoding of the second stimulus s_2^t is error-free, contrary to the first stimulus that is prone to change, with probability ϵ. Furthermore, when s_1 does change, it is replaced by another stimulus, $\hat{s}$ (imposed by the input from the PPC in our network model). Therefore, $\hat{s}$ is drawn from the marginal distribution of bump locations in the PPC, which is similar to the marginal stimulus distribution (see panel B), p_m (see also Sect. 4.2). Depending on the new location of $\hat{s}$, the comparison to s_2 can either lead to an erroneous choice (Bias-, with probability of p_e) or a correct one (Bias+, with probability of $p_c = 1 - p_e$). **(B)** The bump locations in both the WM network (in pink) and the PPC network (in purple) have identical distributions to that of the input stimulus (marginal over s_1 or s_2, shown in gray). **(C)** The distribution of bump locations in PPC (from which replacements $\hat{s}$ are sampled) is overlaid on the stimulus set, and repeated for each value of s_2. For pairs below the diagonal, where $s_1 > s_2$ (red squares), the trial outcome will be an error if the displaced WM bump $\hat{s}$ ends up above the diagonal (blue section of the p_m distribution). The probability to make an error, p_e, scales with the area of the p_m above the diagonal (blue part), which increases as s_1 increases. Similarly, for pairs above the diagonal ($s_1 < s_2$, blue squares), the probability of error scales with the area of p_m that sits below the diagonal, which increases as s_1 decreases. **(D)** The performance of the attractor network as a function of the first stimulus s_1, in red dots for pairs of stimuli where $s_1 > s_2$, and in blue dots for pairs of stimuli where $s_1 < s_2$. The solid lines are fits of the performance of the network using Eq. 9, with ϵ as a free parameter. **(E)** Same data as in the left panel, plotted as fraction classified as $s_1 > s_2$, to illustrate the contraction bias. The colorbar corresponds to fraction classified as $s_1 > s_2$. **(F)** We extend the mathematical model to include one-trial back history effects by assuming that when the first stimulus s_1 does change, it is replaced with s_1^{t-1} (from the previous trial). **(G)** In this case, the performance of the mathematical model naturally yields one-trial back history effects, as quantified through the bias (see text for details). Colored lines correspond to fits of the bias measure for specific pairs of stimuli in the current trial, while the black line is a fit of the mean over all pairs in the current trial, itself shown in black dots.

1.4 Errors are drawn from the marginal distribution of stimuli, giving rise to contraction bias

In order to illustrate the statistical origin of contraction bias in our network model, we consider a mathematical scheme of its performance (Fig. 4 A). In this simple formulation, we assume that the first stimulus to be kept in working memory, s_1^t, is volatile. As a result, in a fraction ϵ of the trials, it is susceptible to replacement with another stimulus $\hat{s}$ (by the input from the PPC network, in our network model). This input from the PPC is sampled from the marginal distribution of bump locations in the PPC, p_m, which is similar to the marginal distribution of stimuli provided in the task (Fig. 4 B). However, this replacement does not always lead to an error, as evidenced by Bias- and Bias+ trials (i.e. those trials in which the performance is affected, negatively and positively, respectively (Fig. 2 B)). For each stimulus pair, the probability to make an error, p_e, is proportional to the area of p_m that is located on the wrong side of the $s_1 = s_2$ diagonal (Fig. 4 C). For instance, for stimulus pairs below the diagonal (Fig. 4 C, red squares) the trial outcome is erroneous only if $\hat{s}$ is displaced above the diagonal (blue part of the distribution). As one can see, the area above the diagonal increases as s_1 increases, giving rise to a gradual increase in error rates (Fig. 4 C). This mathematical model can capture the performance of the attractor network model, as can be seen through the fit of the network performance, with ϵ as a free parameter (see Eq. 9 in Sect. 4.2, Fig. 4 D, E). If we consider that the replacement is not memory-less, i.e. $\hat{s}$ to be one of the recently presented stimuli, e.g. s_2^{t-1} (Fig. 4 F), then the short-term history effects are also recovered (Fig. 4 G).

This simple analysis implies that the contraction bias in the WM network in our model is not the result of the representation of the mean stimulus in the PPC, but is an effect that emerges as a result of the PPC network's sampling dynamics, mostly from recently presented stimuli. Indeed, a "contraction to the mean" hypothesis only provides a general account of which pairs of stimuli should benefit from a better performance and which should suffer, but does not explain the gradual accumulation of errors upon increasing (decreasing) s_1, for pairs below (above) the $s_1 = s_2$ diagonal [11, 31, 12]. Notably, it cannot explain why the performance in trials with pairs of stimuli where s_2 is most distant from the mean stand to benefit the most from it. All together, our model suggests that contraction bias may be a simple consequence of errors occurring at single trials, driven by inputs from the PPC that follow a similar distribution as that of the external input.

1.5 Model predictions

1.5.1 The stimulus distribution impacts the pattern of contraction bias

In our model, errors are determined by the cumulative distribution of stimuli from the theoretical decision boundary $s_1 = s_2$ to the left (right) for pairs of stimuli below (above) the diagonal (Fig. 4 C). This implies that using a stimulus set in which this distribution is skewed makes different predictions for the gradient of performance across different stimulus pairs. A distribution that is symmetric (Fig. 4 C) yields an equal performance for pairs below and above the $s_1 = s_2$ diagonal (blue and red lines) when s_1 is at the mean (as well as the median, given the symmetry of the distribution). A distribution that is skewed, instead, yields an equal performance when s_1 is at the median for both pairs below and above the diagonal. For a negatively skewed distribution, performance is a concave (convex) function of s_1 for pairs of stimuli below (above) the diagonal (Fig. 5 A). Conversely, for a positively skewed distribution, performance is a convex (concave) function of s_1 for pairs of stimuli below (above) the diagonal (Fig. 5 B). For a distribution that is bimodal, the performance as a function of s_1 resembles a saddle, with equal performance for intermediate values of s_1 (Fig. 5 C). These results indicate that although the performance is quantitatively shaped by the exact form of the stimulus distribution, it persists as a monotonic function of s_1 under a wide variety of manipulations of the distributions. This is a result of the property of the cumulative function, and may underlie the ubiquity of contraction bias under different experimental conditions.

1.5.2 A prolonged inter-trial interval improves average performance and reduces bias

If errors are due to the persistence of activity resulting from previous trials, what then, is the effect of the inter-trial interval (ITI)? In our model, a shorter ITI (relative to the default value of $5s$ used in Figs. 2 and 3) results in a worse performance (Fig. 6 A, B, C). The opposite is true for longer ITIs. This change in performance is reflected in reduced biases toward the previous trial (Fig. 6 D and E). In our model, a prolonged ITI allows for a drifting bump to vanish due to two mechanisms. In the first mechanism, if the bump corresponds to a stimulus shown several trials before the pre-ITI trial, it dissipates under the combined effect of adaptation and inhibition resulting from incoming external stimuli. If it corresponds to the stimulus shown in the pre-ITI trial, it eventually reaches one of the two boundaries, where it eventually vanishes due to the free boundary conditions. As a result, the performance improves with increasing ITI and conversely, worsens with a shorter ITI.

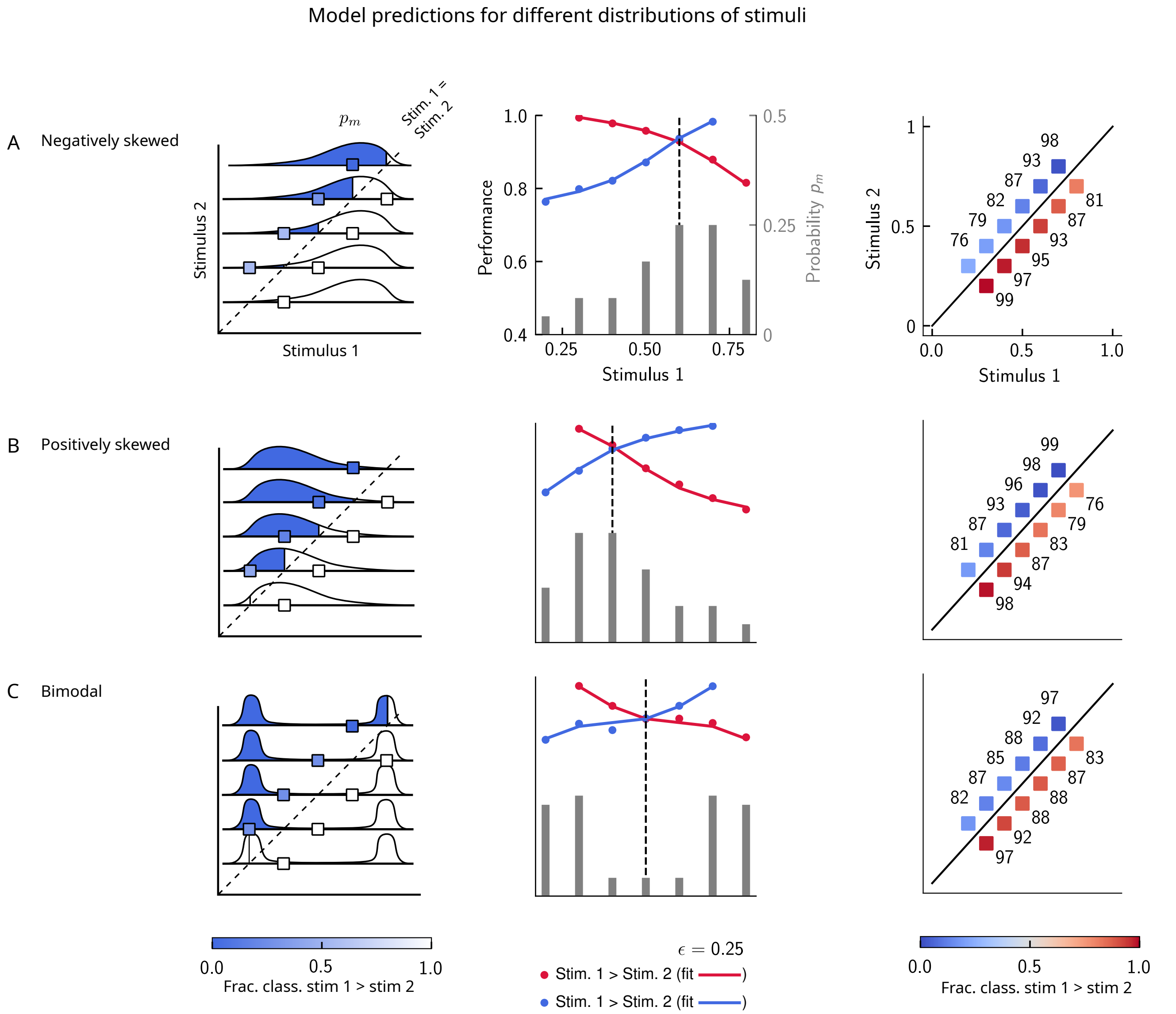

Figure 5: The stimulus distribution impacts the pattern of contraction bias. The model makes different predictions for the performance, depending on the shape of the stimulus distribution. **(A)** Left: Schema of model prediction. Regions shaded in blue correspond to the probability of correct comparison, for stimulus pairs above the diagonal, when replacing s_1 with a random value sampled from the marginal distribution with a probability $\epsilon = 0.25$ (see Fig. 4). If the stimulus distribution is negatively skewed, then performance for pairs of stimuli below (above) the diagonal is concave (convex). The colorbar corresponds to the fraction of trials classified as $s_1 > s_2$. Middle: the probability of different values of s_1 is presented in grey bars (to be read with the right y-axis). The performance of the mathematical model for pairs of stimuli below (above) the diagonal is in red (blue) dots (to be read with the left y-axis). Solid lines correspond to fits with Eq. 9. The value of s_1 for which there is equal performance for pairs of stimuli below and above the diagonal is indicated by the vertical dashed line, coinciding with the median of the distribution. Right: model performance for individual stimulus pairs. **(B)** Similar to A, for a positively skewed distribution. **(C)** Similar to A, for a bimodal distribution.

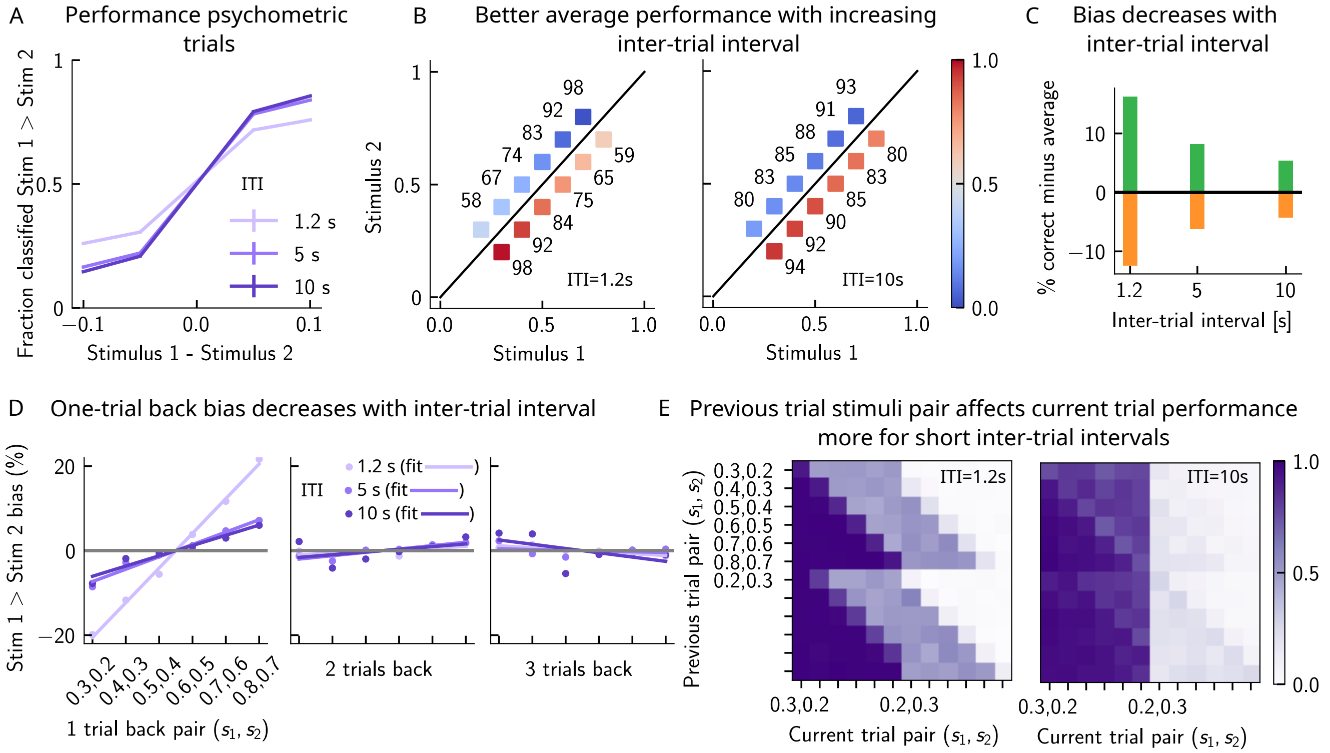

Figure 6: A prolonged inter-trial interval (ITI) improves average performance and reduces bias. **(A)** Performance of the network model for the psychometric stimuli improves with an increasing inter-trial interval. Errorbars (not visible) correspond to the s.e.m. over different simulations. **(B)** The network performance is on average worse for longer ITIs (right panel, $ITI = 10s$), compared to shorter ones (left panel, $ITI = 1.2s$). Colorbar indicates the fraction of trials classified as $s_1 > s_2$. **(C)** Quantifying contraction bias separately for Bias+ trials (green) and Bias- trials (orange) yields a decreasing bias as the inter-trial interval increases. **(D)** Left panel: bias, quantifying the (attractive) effect of the previous trial stimulus pair. Different shades of purple correspond to bias for different values of the ITI, with dots corresponding to simulation values and lines of the same color to linear fits. The one-trial back history effects are attractive: the larger the previous stimulus, the higher the probability of classifying the first stimulus s_1 as large, and vice-versa. The attractive bias is larger for a smaller ITI (light purple, $ITI = 1.2s$), and smaller for a larger ITI (dark purple, $ITI = 10s$). Middle/right panels: same as the left panel, for stimuli extending two and three trials back. **(E)** Performance is affected by the previous stimulus pairs (modulation along the y-axis), more for a short ITI (left, $ITI = 1.2s$) than for a longer ITI (right, $ITI = 10s$). The colorbar corresponds to the fraction classified $s_1 > s_2$. For readability, only some labels are shown.

1.6 The timescale of adaptation in the PPC network can control perceptual biases similar to those observed in dyslexia and autism

In a recent study [8], a similar PWM task with auditory stimuli was studied in human neurotypical (NT), autistic spectrum (ASD) and dyslexic (DYS) subjects. Based on an analysis using a generalized linear model (GLM), a double dissociation between different subject groups was suggested: ASD subjects exhibit a stronger bias towards long-term statistics – compared to NT subjects –, while for DYS subject, a higher bias is present towards short-term statistics.

We investigated our model to see if it is able to show similar phenomenology, and if so, what are the relevant parameters controlling the timescale of the biases in behaviour. We identified the adaptation timescale in the PPC as the parameter that affects the extent of the short-term bias, consistent with previous literature [34], [35]. Calculating the mean bias towards the previous trial stimulus pair (Fig. 7 A), we find that a shorter-than-NT adaptation timescale yields a larger bias towards the previous trial stimulus. Indeed, a shorter timescale for neuronal adaptation implies a faster process for the extinction of the bump in PPC – and the formation of a new bump that remains stable for a few trials – producing a “jumpier” dynamics that leads to a larger number of one-trial back errors. In contrast, increasing this timescale with respect to NT gives rise to a stable bump for a longer time, ultimately yielding a smaller short-term bias. This can be seen in the detailed breakdown of the behavior on the current trial, when conditioned on the stimuli presented at the previous trial (Fig. 7 B, see also Sect. 1.3 for a more detailed explanation of the dynamics). We performed a GLM analysis as in Ref. [8] to the network behavior, with stimuli from four trials back and the mean stimulus as regressors (see Sect. 4.3). This analysis shows that a reduction in the PPC adaptation timescale with respect to NT, produces behavioural changes qualitatively compatible with data from DYS subjects; on the contrary, an increase of this timescale yields results consistent with ASD data (Fig. 7 C).

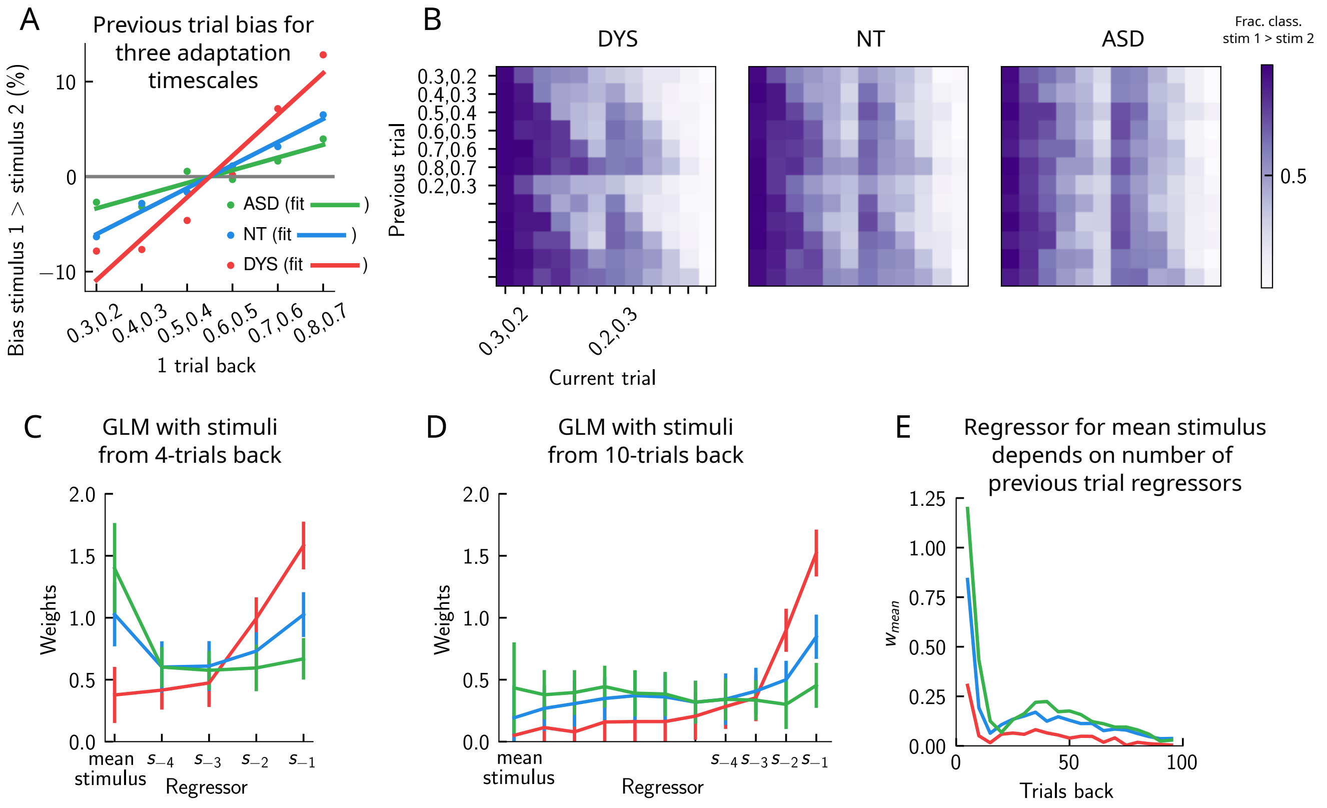

Figure 7: Apparent trade-off between short- and long-term biases, controlled by the timescale of neural adaptation. (A) the bias exerted on the current trial by the previous trial (see main text for how it is computed), for three values of the adaptation timescale that mimic similar behavior to the three cohorts of subjects. **(B)** As in (Fig. 2 D), for three different values of adaptation timescale. The colorbar corresponds to the fraction classified $s_1 > s_2$. **(C)** GLM weights corresponding to the three values of the adaptation parameter marked in (Fig. S5 A), including up to 4 trials back. In a GLM variant incorporating a small number of past trials as regressors, the model yields a high weight for the running mean stimulus regressor. Errorbars correspond to the standard deviation across different simulations. **(D)** Same as in C, but including regressors corresponding to the past 10 trials as well as the running mean stimulus. With a larger number of regressors extending into the past, the model yields a small weight for the running mean stimulus regressor. Errorbars correspond to the standard deviation across different simulations. **(E)** The weight of the running mean stimulus regressor as a function of extending the number of past trial regressors decays upon increasing the number of previous-trial stimulus regressors.

This GLM analysis suggests that dissociable short- and long-term biases may be present in the network behavior. Having access to the full dynamics of the network, we sought to determine how it translates into such dissociable short- and long-term biases. Given that all the behavior arises from the location of the bump on the attractor, we quantified the fraction of trials in which the bump in the WM network, before the onset of the second stimulus, was present in the vicinity of any of the previous trial's stimuli (Fig. S5 B, right panel, and C), as well as the vicinity of the mean over the sensory history (Fig. S5 B, left panel, and C). While the bump location correlated well with the GLM weights corresponding to the previous trial's stimuli regressor (comparing the right panels of Fig. S5 A and B), surprisingly, it did not correlate with the GLM weights corresponding to the mean stimulus regressor (comparing the left panels of Fig. S5 A and B). In fact, we found that the bump was in a location given by the stimuli of the past two trials, as well as the mean over the stimulus history, in a smaller fraction of trials, as the adaptation timescale parameter was made larger (Fig. S5 C).

Given that the weights, after four trials in the past, were still non-zero, we extended the GLM regression by including a larger number of past stimuli as regressors. We found that doing this greatly reduced the weight of the mean stimulus regressor (Fig. 7 C, D and E, see Sect. 4.3 and 4.3 for more details). Therefore, we propose an alternative interpretation of the GLM results given in Ref. [8]. In our model, the increased (reduced) weight for long-term mean in the ASD (DYS) subjects can be explained as an effect of a larger (smaller) window in time of short-term biases, without invoking a double dissociation mechanism (Fig. 7 D and E). In Sect. 4.3, we provide a mathematical argument for this, which is empirically shown by including a large number of individual stimuli from previous trials in the regression analysis.

2 Discussion

Contraction bias is an effect emerging in working memory tasks, where the averaged behavior suggests that the magnitude of the item held in memory is perceived as larger when it is "small" and, vice-versa, it is perceived as smaller when it is "large" [2–5, 7, 9, 10]. Recently, it has been shown that contraction bias as well as short-term history-dependent effects occur in an auditory delayed comparison task in rats and humans: the comparison performance in given trial, depends on the stimuli shown in preceding trials (up to three trials back) [12]. This study raises the question: does contraction bias occur independently of short-term history effects, or does it emerge as a result of the latter? Our model provides support for the second possibility, offering a parsimonious account of how contraction bias emerges [36].

Our WM and PPC networks, despite having different parameters, are both shown to encode information about the marginal distribution of the stimuli (Fig. 4 B). However, despite having similar activity distributions to that of the external stimuli, they have different memory properties, due to the different timescales with which they process incoming stimuli. The putative WM network, from which information to solve the task is read-out, receives additional input from the PPC network. The PPC, in turn, is modelled as integrating inputs slower relative to the WM network, and is also endowed with firing rate adaptation. As a result, the activity bump in the PPC drifts slowly, until it is extinguished due to the combined effect of adaptation and global inhibition due to external inputs. Once this occurs, the upcoming external input forms a new bump, that can remain stable for some trials. The newly formed bump in PPC has the highest chance to impact the WM activity bump and give rise to short-term history effects.

However, short-term history effects do not need to be invoked in order to explain contraction bias. As long as errors are made following random samples from a distribution in the same range as that of the stimuli, contraction bias should be observed. Indeed, when we manipulated the parameters of the PPC network in such a way that short-term history effects were eliminated (by removing the firing-rate adaptation), contraction bias persisted.

As a result, our model suggests that contraction bias may not simply be given by a regression towards the mean of the stimuli during the inter-stimulus interval [37, 38], but brought about by a richer dynamics [6], more in line with the idea of random sampling [39]. The model makes predictions as to how the pattern of errors may change when the distribution of stimuli is manipulated, either at the level of the presented stimuli or through the network dynamics. In support of this, in a recent tactile categorization study [40], where rats were trained to categorize tactile stimuli according to a boundary set by the experimenter, the authors have shown that rats set their decision boundary according to the the statistical structure of the stimulus set to which they are exposed.

We note that in our model, the stimulus distribution is not explicitly learned (but see [41]). Instead, since the PPC dynamics slowly follows the input, its marginal distribution of activity bump will be similar to the marginal distribution of the external input. Support for this idea comes from Ref. [40], where the authors used different stimulus ranges across different sessions and noted that rats initiated each session without any observable residual influence of the previous session's range/boundary on the current session.

Another framework that has been commonly used to account for contraction bias is Bayesian inference [39, 42]. In such a framework, the abstract mathematical model that we present, would be recovered by Ref. [7] in the limit of a very broad likelihood for the first stimulus and a very narrow one for the second stimulus, and where the prior for the first stimulus is replaced by the distribution of $\hat{s}$, following the model in Fig. 4 A (see Sect. 4.2 for details). However, our model of decision making is conceptually different: the subject does not have access to the full prior distribution, as in the Bayesian framework (where it can be used to make optimal decisions), but only to *samples* of the prior.

The interpretation of the contraction bias presented in this work pertains to the information processing only –i.e. the storage of recent stimuli. Alternative statistical models, e.g. in Ref. [12], assume an uncertainty at the decision-making stage –i.e. include a sigmoidal response function that partially accounts for the errors. A description of these effects is beyond the scope of our work: although these are certainly important, we show that they are not necessary in order to explain the occurrence of sensory history biases. Equivalently, one could also include volatility in the representation of s_2, prior to decision making.

Although contraction bias is robustly found in different contexts, surprisingly similar tasks, such as perceptual estimation tasks, sometimes highlight opposite effects, i.e. repulsive effects [43, 44, 40]. Sensory adaptation has long been evoked to explain such findings [45, 43, 46], mostly studied in the context of vision, and shown to underly some illusion effects. Yet other studies have found both effects in the same experiment. For example, in a study of visual orientation estimation, the authors show that attraction and repulsion have different timescales; while perceptual decisions about orientation are attracted towards recently perceived stimuli (timescale of a few seconds), they are repelled from stimuli that were shown further back in time (timescale of a few minutes), consistent with neuronal adaptation [47]. Moreover, they demonstrate that the long-term repulsive bias is spatially specific, in line with sensory adaptation [48–50] and in contrast to short-term attractive serial dependence [44]. Several studies propose that distinct mechanisms may be at the basis of the seemingly contradictory effects. In a study using visual backward masking [51], the authors show that suppressing high-level modulatory signals on early cortical activity eliminates the attractive effect of previous trials. Moreover, they observe robust repulsive effects, in line with perceptual adaptation, after only 50 ms of stimulation. In another study, Pitts et al. observed repulsive effects when strong neural adaptation was present, and priming effects in conditions where adaptation was reduced and instead, attention was recruited [52]. These findings suggest coexistence and interaction of bottom-up and top-down processes, where local network mechanisms give rise to repulsive effects, and top-down cortical inputs modulate the attractive effects [53, 54]. The short-term attractive effects in our model of WM, arising as a result of the PPC activity are consistent with these findings.

In yet another spatial delayed response task study, the delay between stimulus and response was parametrically varied in order to explore the serial dependence from the moment the stimulus was perceived to the moment it was held in post-perceptual visual working memory. The authors found evidence of adaptation in the behavioral responses made immediately after viewing a stimulus, but no evidence for attractive serial dependence. However, as the delay period was increased, attractive serial dependence became apparent [52], also consistent with more recent literature [55]. These results, although from a different modality and task, are consistent with our model; indeed, we find that errors occur during the inter-stimulus interval, once inputs from the PPC become strong enough to disrupt the stimulus held in WM. Also consistent with these results, in our model, contraction bias increases upon increasing the inter-stimulus interval.

Our model assumes that the stimulus is held in working memory through the persistent activity of neurons, building on the discovery of persistent selective activity in a number of cortical areas, including the prefrontal cortex (PFC), during the delay interval [56–62]. To explain this finding, we have used the attractor framework, in which recurrently connected neurons mutually excite one another to form reverberation of activity within populations of neurons coding for a given stimulus [63–65]. However, subsequent research has shown that persistent activity is not always present during the delay period and that the activity of neurons displays far more heterogeneity than previously thought [66]. It has been proposed that short-term synaptic facilitation can dynamically operate to bring a WM network across a phase transition from a silent to a persistently active state [67–69].

Finally, while inputs from the PPC are detrimental to the performance in the context of an task designed in a laboratory setting, this may not be the case in more natural environments. For example, it has been suggested that integrating information over time serves to preserve perceptual continuity in the presence of noisy and discontinuous inputs [14]. Moreover, this continuity of perception may be necessary in order to solve more complex tasks or make decisions, particularly in a non-stationary environment, or in a noisy environment. In our case, no information about the stimulus history extending multiple trials back is necessary to solve the task, a condition that may not always be fulfilled in natural environments.

3 Methods

3.1 The model

Our model is composed of two populations of N neurons, representing the PPC network and the putative WM network. We consider that each population is organized as a continuous line attractor, with recurrent connectivity described by an interaction matrix J_{ij}, whose entries represent the strength of the interaction between neuron i and j. The activation function of the neurons is a logistic function, i.e. the output r_i of neuron i, given the input h_i, is

$$r_i = \frac{1}{1 + e^{-\beta h_i}} \tag{1}$$

where β is the neuronal gain. The variables r_i take continuous values between 0 and 1, and represent the firing rates of the neurons. The input h_i to a neuron is given by

$$\tau \frac{dh_i}{dt} = \sum_{j(\neq i)} J_{ij} r_j + I_i^{\text{ext}} \tag{2}$$

where τ is the timescale for the integration of inputs. In the first term on the right hand side, $J_{ij}r_j$ represents the input to neuron i from neuron j, and I_i^{ext} corresponds to the external inputs. The recurrent connections are given by

$$J_{ij} = \frac{1}{d_0}(K_{ij} - J_0)\,, \tag{3}$$

with

$$K_{ij} = J_e\, e^{-\frac{|x_i - x_j|}{d_0}}\,. \tag{4}$$

The interaction kernel, K, is assumed to be the result of a time-averaged Hebbian plasticity rule: neurons with nearby firing fields will fire concurrently and strengthen their connections, while firing fields far apart will produce weak interactions [70]. Neuron i is associated with the firing field $x_i = i/N$. The form of K expresses a connectivity between neurons i and j that is exponentially decreasing with the distance between their respective firing fields, proportional to $|i - j|$; the exponential rate of decrease is set by the constant d_0, i.e. the typical range of interaction. This constant also sets the amplitude of the the kernel, in such a way that its integral over x equals one. The strength of the excitatory weights is set by J_e; the normalization of K, together with the sigmoid activation function saturating to 1, implies that J_e is also the maximum possible input received by any neuron due to the recurrent connections. The constant J_0, instead, contributes to a linear global inhibition term. Its value needs to be chosen depending on J_e and d_0, so that the balance between excitatory and inhibitory inputs ensures that the activity remains localized along the attractor, i.e. it does not either vanish or equal 1 everywhere; together, these three constants set the width of the bump of activity.

The two networks in our model are coupled through excitatory connections from the PPC to the WM network. Therefore, we introduce two equations analogous to Eq. (2), one for each network. The coupling between the two will enter as a firing-rate dependent input, in addition to I^{ext}. The dynamics of the input to a neuron in the WM network writes

$$\tau_h^W \frac{dh_i^W}{dt} + h_i^W = \sum_{j \in j^W} J_{ij} r_j^W + J^{P \to W} r_i^P + I_i^{\text{ext}}\,, \tag{5}$$

where τ_h^W is the timescale for the integration of inputs in the WM network. The first term in the r.h.s corresponds to inputs from recurrent connections within the WM network. The second term, corresponds to inputs from the PPC network. Finally, the last term corresponds to the external inputs used to give stimuli to the network. Similarly, for the PPC network we have

$$\tau_h^P \frac{dh_i^P}{dt} + h_i^P = \sum_{j \in j^P} J_{ij} r_j^P - \theta_i^P + I_i^{\text{ext}}\,, \tag{6}$$

where τ_h^P is the timescale for the integration of inputs in the PPC network; importantly, we set this to be longer than the analogous quantity for the WM network, $\tau_h^W < \tau_h^P$ (see Tab. 1). The first and third terms in the r.h.s are analogous to the corresponding ones for the WM network: inputs from within the network and from the stimuli. The second term instead, corresponds to adaptive thresholds with a dynamics specified by

$$\tau_\theta^P \frac{d\theta_i^P}{dt} + \theta_i^P = D^P r_i^P \tag{7}$$

modelling neuronal adaptation, where τ_θ^P and D^P set its timescale and its amplitude. We are interested in the condition where the timescale of the evolution of the input current is much smaller relative to that of the adaptation ($\tau_h^P \ll \tau_\theta^P$). For a constant τ_θ^P, we find that depending on the value of D^P, the bump of activity shows different behaviors. For low values of D^P, the bump remains relatively stable (Fig. S1 C (1)). Upon increasing D^P, the bump gradually starts to drift (Fig. S1 C (2-3)). Upon increasing D^P even further, a phase transition leads to an abrupt dissipation of the bump (Fig. S1 C (4)).

Note that, while the transition from bump stability to drift occurs gradually, the transition from drift to dissipation is abrupt. This abruptness in the transition from the drift to the dissipation regime may imply that only one of the two behaviors is possible in our model of the PPC (Sect. 1.3). In fact, our network model of the PPC operates in the "drift" regime ($\tau_\theta^P = 7.5$, $D^P = 0.3$). However, we also observe dissipation of the bump. This occurs due to the inputs from incoming external stimuli, that affect the bump via the global inhibition in the model (Fig. S1 A). Therefore external stimuli can allow the network to temporarily cross the sharp drift/dissipation boundary shown in Fig. S1 B. As a result, the combined effect of adaptation, together with external inputs and global inhibition result in the drift/jump dynamics described in the main text.

Finally, both networks have a linear geometry with free boundary conditions, i.e. no condition is imposed on the profile activity at neuron 1 or N.

3.2 Simulation

We performed all the simulations using custom Python code. Differential equations were numerically integrated with a time step of $dt = 0.001$ using the forward Euler method. The activity of neurons in both circuits were initialized to $r = 0$. Each stimulus was presented for 400 ms. A stimulus is introduced as a "box" of unit amplitude and of width $2\,\delta s$ around s in stimulus space: in a network with N neurons, the stimulus is given by setting $I_i^{\text{ext}} = 1$ in Eq. 5 for neurons with index i within $(s \pm \delta s) \times N$, and $I_i^{\text{ext}} = 0$ for all the others. Only the activity in the WM network was used to assess performance. To do that, the activity vector was recorded at two time-points: 200 ms before and after the onset of the second stimulus s_2. Then, the neurons with the maximal activity were identified at both time-points, and compared to make a decision. This procedure was done for 50 different simulations with 1000 consecutive trials in each, with a fixed inter-trial interval separating two consecutive trials, fixed to 5 seconds. The inter-stimulus intervals were set according to two different experimental designs, as explained below.

3.2.1 Interleaved design

As in the study in Ref. [12], an inter-stimulus interval of either 2, 6 or 10 seconds was randomly selected. The delay interval is defined as the time elapsed from the end of the first stimulus to the beginning of the second stimulus. This procedure was used to produce Figs. 1, 2, 3, 6, S2, S3.

3.2.2 Block design

In order to provide a comparison to the interleaved design, but also to simulate the design in Ref. [8], we also ran simulations with a block design, where the inter-stimulus intervals were kept fixed throughout the trials. Other than this, the procedure and parameters used were exactly the same as in the interleaved case. This procedure was used to produce Figs. S4 and S5.

4 Supplementary Material

4.1 Computing bump location

In order to check whether the bump is in a target location (Figs. 3 B, S2 B, and S3 D), we check whether the position of the neuron with the maximal firing rate is within a distance of ±5% of the length of the whole line attractor from the target location (Figs. 3 A, S2 A and S3 C). In these figures, we compare the probability that, in a given trial, the activity of the WM network is localized around one of the previous stimulus (estimated from the simulation of the dynamics, histograms) with the probability of this happening due to chance (horizontal dashed line). Here we detail the calculation of the chance probability. In general, if we have two discrete independent random variables, $\hat{X}$ and $\hat{Y}$, with probability distributions p_X and p_Y, the probability of them having the same value is

$$\text{Prob}\{\hat{X} = \hat{Y}\} = \sum_{i,j} \underbrace{\text{Prob}\{\hat{X} = x_i\}}_{p_X^i} \underbrace{\text{Prob}\{\hat{Y} = y_j\}}_{p_Y^j} \mathbb{I}(x_i = y_j)$$

where i, j are the indices for different values of the two random variables and $\mathbb{I}(x_i = y_j)$ equals 1 where $x_i = x_j$ and 0 otherwise. If the two random variables are identically distributed, the above expression writes

$$\text{Prob}\{\hat{X} = \hat{Y}\} = \sum_{i,j} p_X^i \, p_Y^j \, \delta_{i,j} = \sum_i \left(p_X^i\right)^2$$

In our case, the two identically distributed random variables are "bump location at the current trial" and the "target bump location" (that are s_1^{t-2}, s_2^{t-2}, s_1^{t-1}, s_2^{t-1} and $\langle s \rangle$). With the exception of the mean stimulus $\langle s \rangle$, all the other variables are identically distributed, with probability p_m (that is the marginal distribution over s_1 or s_2). We note that the bump location in the WM network follows a very similar distribution to p_m (Fig. 4 B). Then, we compute the chance probability with the above relationship, where $p_X \equiv p_m$. For the mean stimulus, instead, we have a probability which is simply equal to 1 for $s = 0.5$ and 0 elsewhere; therefore, the chance probability for the bump location to be at the mean stimulus, then is $p_m(0.5)$.

The excess probability (with respect to chance) for the bump location to equal one of the previous stimuli gives a measure of the correlation between these two; in other terms, of the amount of information retained by the network about previous stimuli.

4.2 The probability to make errors is proportional to the cumulative distribution of the stimuli, giving rise to contraction bias

In order to illustrate the statistical origin of contraction bias consistent with our network model, we consider a simplified mathematical model of its performance (Fig. 4 A). By definition of the delayed comparison task, the optimal decision maker produces a label y equal to 0 if $s_1^t < s_2^t$, and 1 if $s_1^t > s_2^t$; the impossible cases $s_1^t = s_2^t$ are excluded from the set of stimuli, but would produce a label which is either 0 or 1 with 50% probability. That is

$$y(s_1, s_2) = \begin{cases} 1 & \text{if } s_1 > s_2 \\ 0 & \text{if } s_1 < s_2 \\ \text{Bernoulli}(1/2) & \text{if } s_1 = s_2 \end{cases} \tag{8}$$

In this simplified scheme, at each trial t, the two stimuli s_1^t and s_2^t are perfectly perceived with a finite probability $1 - \epsilon$, with $\epsilon < 1$. Under the assumption that the decision maker behaves optimally based on the perceived stimuli, a correct perception would necessarily lead to the correct label. However, with probability ϵ, the first stimulus is randomly selected from a buffer of stimuli, i.e. is replaced by a random variable $\hat{s}_1$ that has a probability distribution p_m^t.

The probability distribution p_m^t is the statistics of previously shown stimuli. The information about the previous stimulus is given by the activity of the "slower" PPC network. As shown above, after the presentation of the first stimulus of the trial, the bump of activity is seen to jump to the position encoding one of the previously presented stimuli, s_2^{t-1}, s_1^{t-1}, s_2^{t-2}, etc. with decreasing probability (Fig. 3 C). Therefore, in calculating the performance in the task, we can take p_m^t to be the marginal distribution of the stimulus s_1 or s_2 across trials, as in the histogram (Fig. 4 B).

The probability of a misclassification is then given by the probability that, given the pair (s_1^t, s_2^t), at trial t,

1) the first stimulus is replaced by a random value, which happens with probability ϵ, and

2) the value of $\hat{s}_1$ replaced is larger than s_2^t when s_1^t is smaller and viceversa (Fig. 4 C).

In summary, the probability of an error at trial t is given by

$$\text{Prob}\left\{\text{error} \;\middle|\; s_1^t = s_1,\, s_2^t = s_2\right\} = \epsilon \cdot \begin{cases} p_m^t(s_2)/2 + \sum_{s<s_2} p_m^t(s) & \text{if } s_1 > s_2 \;, \\ p_m^t(s_2)/2 + \sum_{s>s_2} p_m^t(s) & \text{if } s_1 < s_2 \;. \end{cases} \tag{9}$$

4.3 Generalized Linear Model (GLM)

GLM as in Lieder et al.

Similarly to Ref. [8], we performed a multivariate logistic regression (an instance of generalized linear model, GLM) to the output of the network in the delayed discrimination task with recent stimuli values as covariates:

$$P(\text{“}s_1^t > s_2^t\text{”}) = \sigma\Big(\alpha\,(s_2^t - s_1^t) - \sum_{i=1}^{h} w_i\,(\overline{s^{t-i}} - s_1^t) - w_{mean}\,(\langle s\rangle - s_1^t)\Big) \tag{10}$$

where σ is the sigmoidal function $\sigma(z) = 1/(1 + e^{-z})$, $\overline{s^\tau} = (s_1^\tau + s_2^\tau)/2$ is the mean of the stimuli presented at trial τ, h is the number of "history" terms in the regression, and $\langle s\rangle$ is the mean of the stimuli within and across trials up to the current one. As in Ref. [8], we choose $h = 4$, i.e. we include in the short-term history, the four trials prior to the current one. The first term in Eq. (10), with weight α, controls the slope of the psychometric curve. The remaining terms, combined linearly with weights w, contribute to biases expressing the long and short-term memory. In Ref. [8], it is shown that subjects on the autistic syndrome (ASD) conserve the higher long-term weights, w_{mean}, while losing the short-term weights expressed by neurotypical (NT) subjects. In contrast, dyslexic (DYS) subjects conserve a higher bias from the recent stimuli, w_1, while losing the higher long-term weights, also expressed by neurotypical subjects.

In order to gain insight into this regression model in terms of our network, we also performed a linear regression of the bump of activity just before the onset of the second stimulus, denoted $\hat{s}_1^t$, versus the same variables:

$$\hat{s}_1^t = s_1^t + \sum_{i=1}^{h} w_i \left(\overline{s^{t-i}} - s_1^t\right) + w_{mean} \left(\langle s \rangle - s_1^t\right) \tag{11}$$

In this case, we see that the weights w in the linear regression for $\hat{s}_1^t$ have the same qualitative behaviour as the weights for the bias term in the GLM regression for the performance (not shown). This is expected, since the decision-making rule in the network –based on the bump location just before and during the second stimulus, $\hat{s}_1$ and $\hat{s}_2 \simeq s_2^t$, respectively– is deterministic, following $P(\text{“}s_1^t > s_2^t\text{”}) = \Theta(s_2^t - \hat{s}_1^t)$. Therefore, the bias term in the GLM performed in Ref. [8], Eq. (10), corresponds to the displacement of the bump location $\hat{s}_1^t$ with respect to the actual stimulus s_1^t, modelled to be linearly dependent on the displacement of previous stimuli from s_1^t.

Regression model with infinite history

In the regression formulas in Eqs. (10) and (11), it is possible to give an interpretation of the parameter w_{mean}, that is the weight of the contribution from the covariate corresponding to the mean of the past stimuli. Let us consider two regression models, one in which, in addition to a regressor corresponding to the mean stimulus, regressors corresponding to the stimulus history are included up to trial h, and another in which $h = \infty$, i.e. infinitely many past stimuli are included as regressors. In this case, Eq. (11) rewrites

$$\hat{s}_1^t = s_1^t + \sum_{i=1}^{\infty} w_i \left(\overline{s^{t-i}} - s_1^t\right). \tag{12}$$

If we assume that the weights obtained from the regression have roughly an exponential dependence on time (Fig. S5 D), we can write

$$w_i = \gamma\, w_{i-1} = \gamma^i\, w_0\,. \tag{13}$$

By equating Eqs. (11) and (12), we would find that

$$\begin{aligned} w_{mean} \left(\langle s \rangle - s_1^t\right) &= \sum_{i=h+1}^{\infty} w_i \left(\overline{s^{t-i}} - s_1^t\right) \\ &= w_{i+1} \sum_{j=0}^{\infty} \gamma^j \left(\overline{s^{t-(h+1+j)}} - s_1^t\right) \\ &= \frac{w_{h+1}}{1-\gamma} \left(\langle s \rangle_\gamma - s_1^t\right) \end{aligned} \tag{14}$$

where

$$\langle s \rangle_\gamma = \sum_{j=0}^{\infty} g_j\, \overline{s^{t-h-1-j}} \tag{15}$$

that is an average over the geometric distribution $g_j = (1-\gamma)\,\gamma^j$, from time $t-(h+1)$ backward. Since for γ large enough we have $\langle s \rangle_\gamma = \langle s \rangle$, we can identify

$$w_{mean} \propto \frac{w_{h+1}}{1-\gamma}\,. \tag{16}$$

This derivation indicates that the magnitude w_{mean} in the infinite history model, given by Eq. (11), is a function of the discount factor γ as well as the weight of the first trial left out from the finite-history regression (w_{h+1}). A higher γ value, i.e. a longer timescale for damping of the weights extending into the stimulus history, yields a higher w_{mean}. We can obtain γ for each condition (NT, ASD and DYS) by fitting the weights obtained as a function of trials extending into history (Fig. S5 D). As predicted by Eq. (16), a larger window for short-term history effects (as in the ASD case relative to NT) yields a larger weight for the covariate corresponding to the mean stimulus. Finally, Eq. (16) also predicts that w_{mean} is proportional to w_{h+1}, the number of trials back we consider in the regression, h, implying that the number of covariates that we choose to include in the model may greatly affect the results. Both of these predictions are corroborated by plotting directly the value of w_{mean} obtained from the regression (Fig. S5 E).

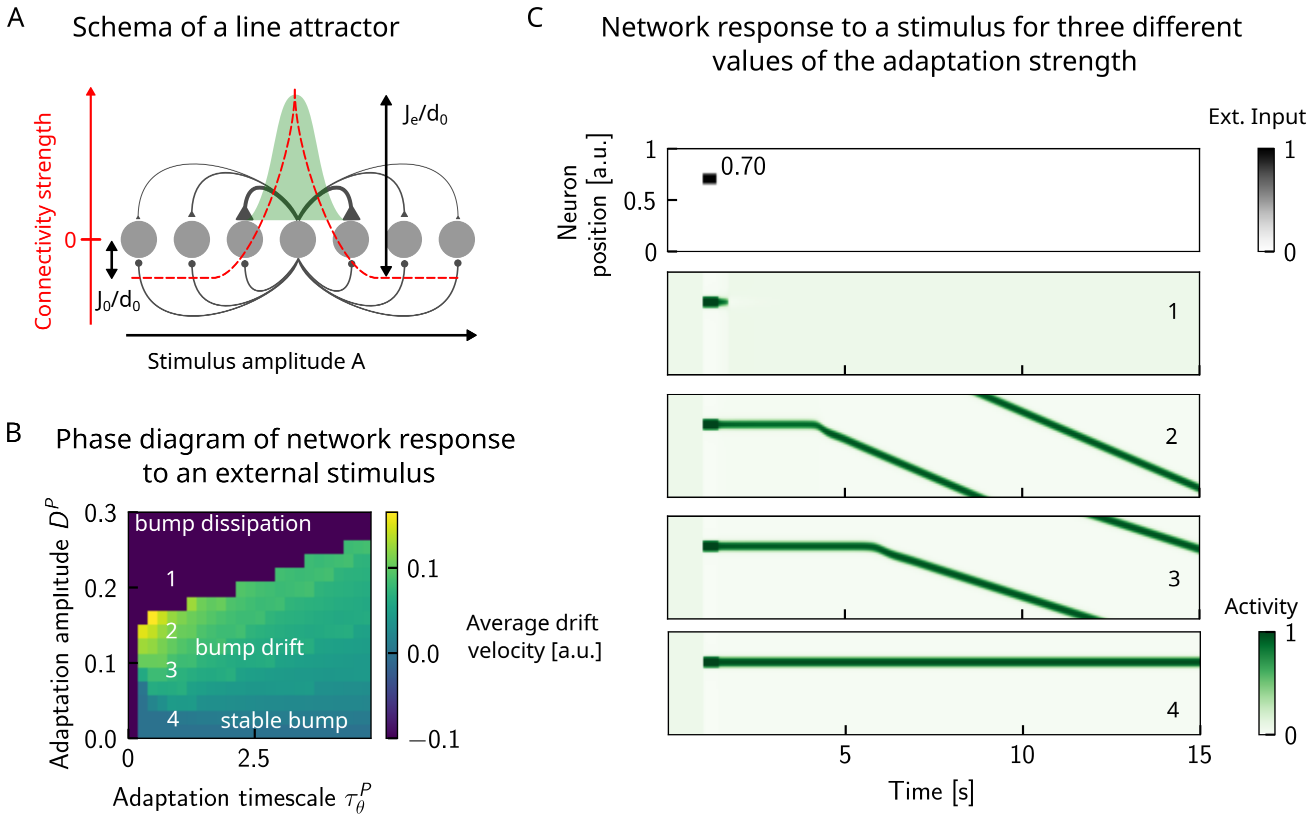

Figure S1: Dynamics of responses in a 1D continuous attractor network, in the presence of adaptation. (A) We study a one dimensional line attractor in which neurons code for a stimulus feature that varies along a physical dimension, such as amplitude of an auditory stimulus. The connections between pairs of neurons is a decreasing, symmetric function of the distance between their preferred firing locations, allowing for a bump of activity to form and self-sustain when sufficient input is given to the network. However, this self-sustaining activity may be disrupted if neuronal adaptation is present. In particular, drifting dynamics may be observed. **(B)** Left: phase diagram of the average drift velocity as a function of the adaptation timescale τ_θ^P and amplitude D^P. The average drift velocity is simply computed as the distance travelled by the center of the bump in a duration of 50 seconds. Color codes for the average drift velocity (a.u.). Numbers indicate four points for which sample dynamics are shown in (C). **(C)** We observe three main phases: in the first, the activity bump is stable when no or little neuronal adaptation is present (point 4). Larger values of neural adaptation induce drift of the activity bump; the average drift velocity increases upon increasing the neural adaptation (points 2 and 3). Finally, increasing it even further leads to the dissipation of the activity bump (point 1). The boundary between the drift and dissipation phases is abrupt. In these simulations, periodic boundary conditions have been used in order to compute the average average drift velocity over longer durations.

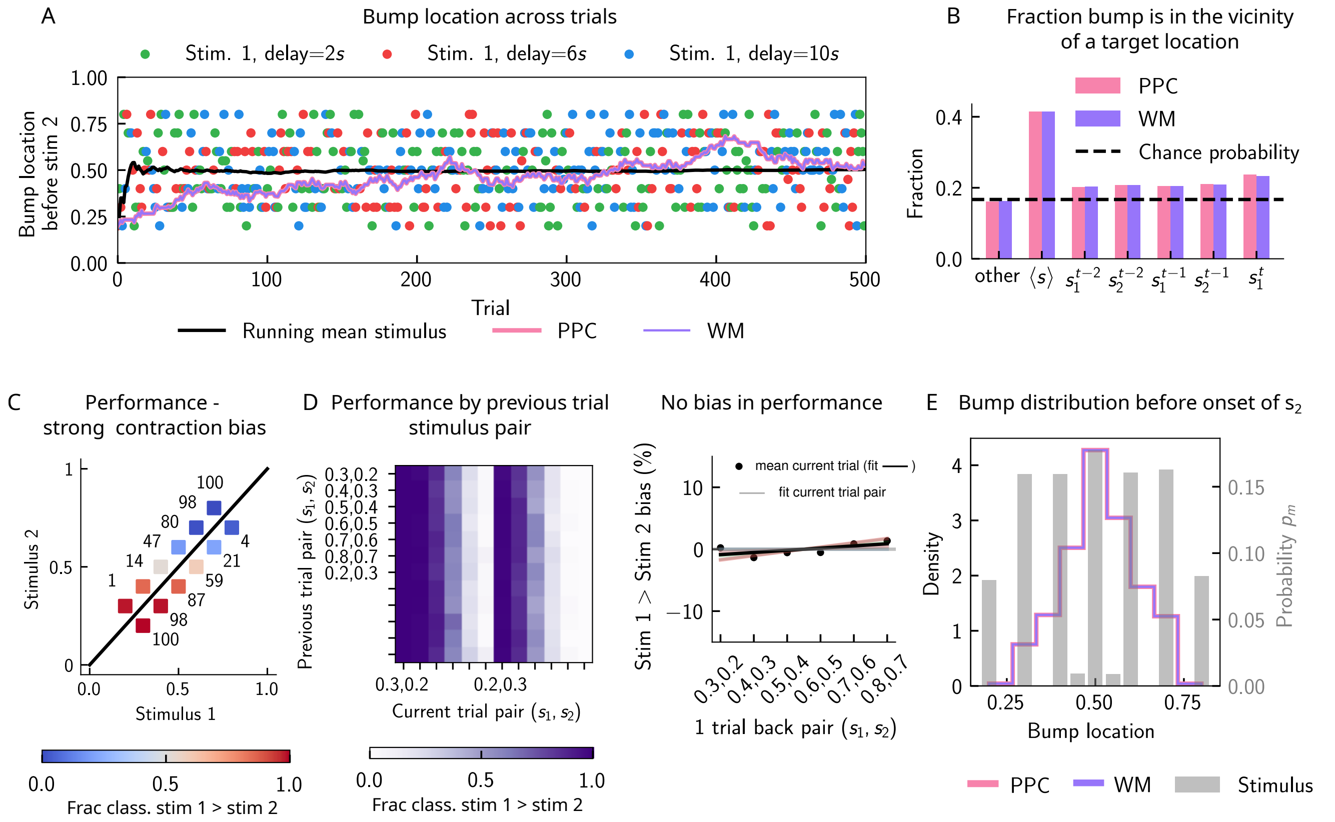

Figure S2: The role of neural adaptation in short-term history biases. In order to better understand the network mechanisms that give rise to short-term history effects, we removed neural adaptation in the PPC network and assessed the performance in the WM network. **(A)** As in (Fig. 3 B). We track the location of the bump, in the PPC (pink), and in the WM network (purple) before the onset of the second stimulus (the pink curve cannot be seen as the purple curve goes perfectly on top). In this case, the displacement of the bump of activity is smooth and new sensory stimuli (colored dots) induce only a minimal shift in the location of the bump. This behavior is to be contrasted with the case in which there is adaptation in the PPC network, inducing jumps in the bump location (Fig. 3 A). An additional effect of no neural adaptation is that the activity in the PPC network, completely overrides the activity in the WM network. **(B)** As in (Fig. 3 C). We compute the fraction of times the bump is in a given location, current trial (s_1^t), four preceding trials (s_1^{t-2} to s_2^{t-1}), the running mean stimulus, or all other locations (overlapping sets). In this case, in the majority of the trials, the bump is either at the running mean stimulus, or any other location. The fraction of trials in which it is in the position of the four previous stimuli roughly corresponds to chance occurrence (dashed black lines), with only a minor increase for the current stimulus. **(C)** As in (Fig. 2 C). In this setting, the performance expresses a very strong contraction bias, and it is as if the decision boundary is orthogonal to the optimal decision boundary. Color codes for fraction of trials in which a $s_1 > s_2$ classification is made. **(D)** As in (Fig. 2 D) Left: the performance conditioned on the previous trial stimulus pair does not exhibit any previous-trial history dependence (vertical modulation). Colorbar corresponds to the fraction classified as $s_1 > s_2$. Right: this can also be expressed through the bias measure (see main text for how it is computed). Colored lines correspond to current trial pairs, the black dots to the mean over all current trial pairs, and the black line to its linear fit. **(E)** As in (Fig. 4 B). Marginal distribution of the bump location in both networks (pink for PPC, purple for WM) before the onset of s_2 is more peaked than the marginal distribution of the stimuli (gray), as a result of the absence of “jumps”.

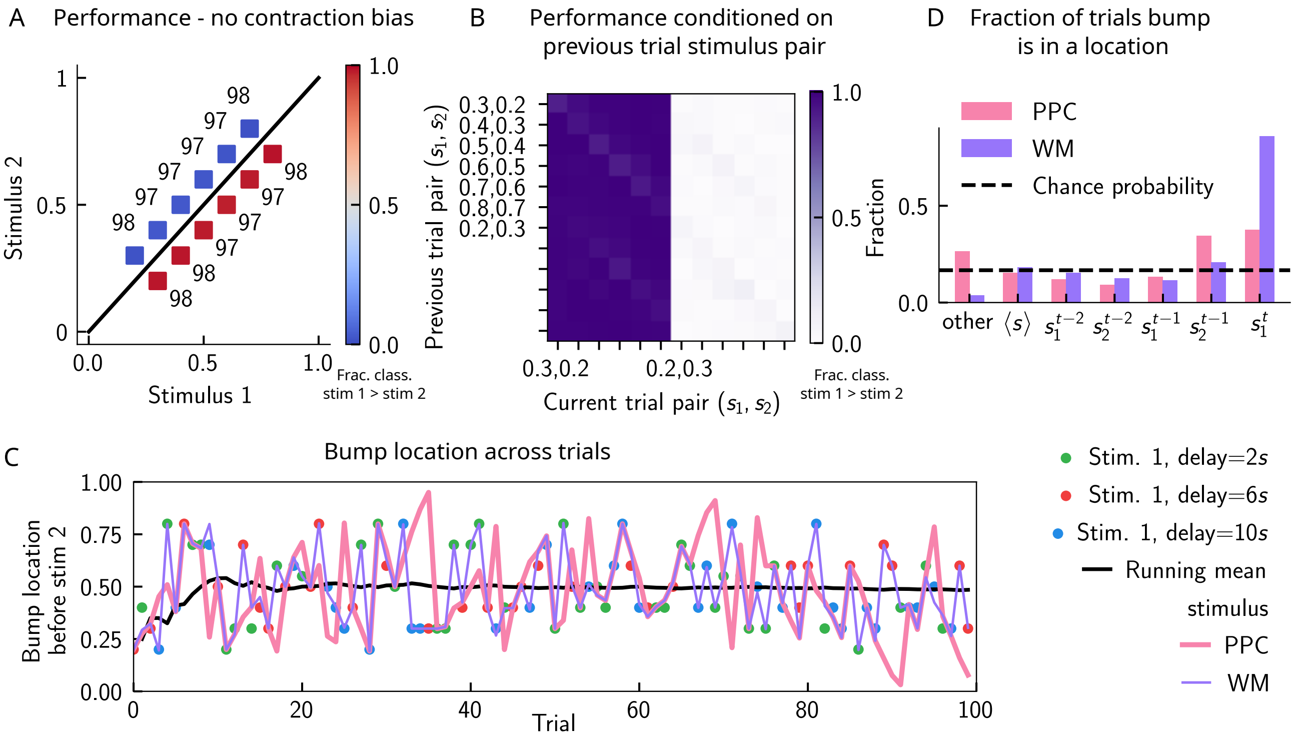

Figure S3: Inactivating the inputs from the PPC network improves performance, in line with experimental findings. **(A)** As in (Fig. 2 C). The performance of the network when the strength of the inputs from the PPC to the WM network is weakened (modelling the optogenetic inactivation of the PPC) is dramatically improved, and contraction bias is virtually eliminated. The colorbar corresponds to the fraction classified as $s_1 > s_2$. **(B)** As in (Fig. 2 D). The performance for each stimulus pair in the current trial is improved and no modulation by the previous stimulus pairs can be observed. The colorbar corresponds to the fraction classified as $s_1 > s_2$. **(C)** As in (Fig. 3 B). This improvement of the performance can be traced back to how well the activity bump in the WM network (in purple), before the onset of the second stimulus s_2, tracks the first stimulus s_1 (shown in colored dots, each corresponding to a different value of the inter-stimulus delay interval). Relative to the case in which inputs from the PPC are intact (Fig. 3 A), it can be seen that the location of the bump tracks the first stimulus with high fidelity. The activity in the PPC (in pink), instead, is identical to that shown previously (Fig. 5 A), as all the other parameters are kept constant. **(D)** As in (Fig. 2 C). The bump location can be quantified not only for the stimulus s_1 of the current trial (colored dots, each color corresponding to a given delay interval), but for the four preceding stimuli from the two previous trials (from s_2^{t-1} back to s_1^{t-2}). With weaker inputs from the PPC (pink), the WM (purple) function of the circuit is disrupted less frequently, and in the majority of the trials, the bump of activity corresponds to the first stimulus s_1^t.

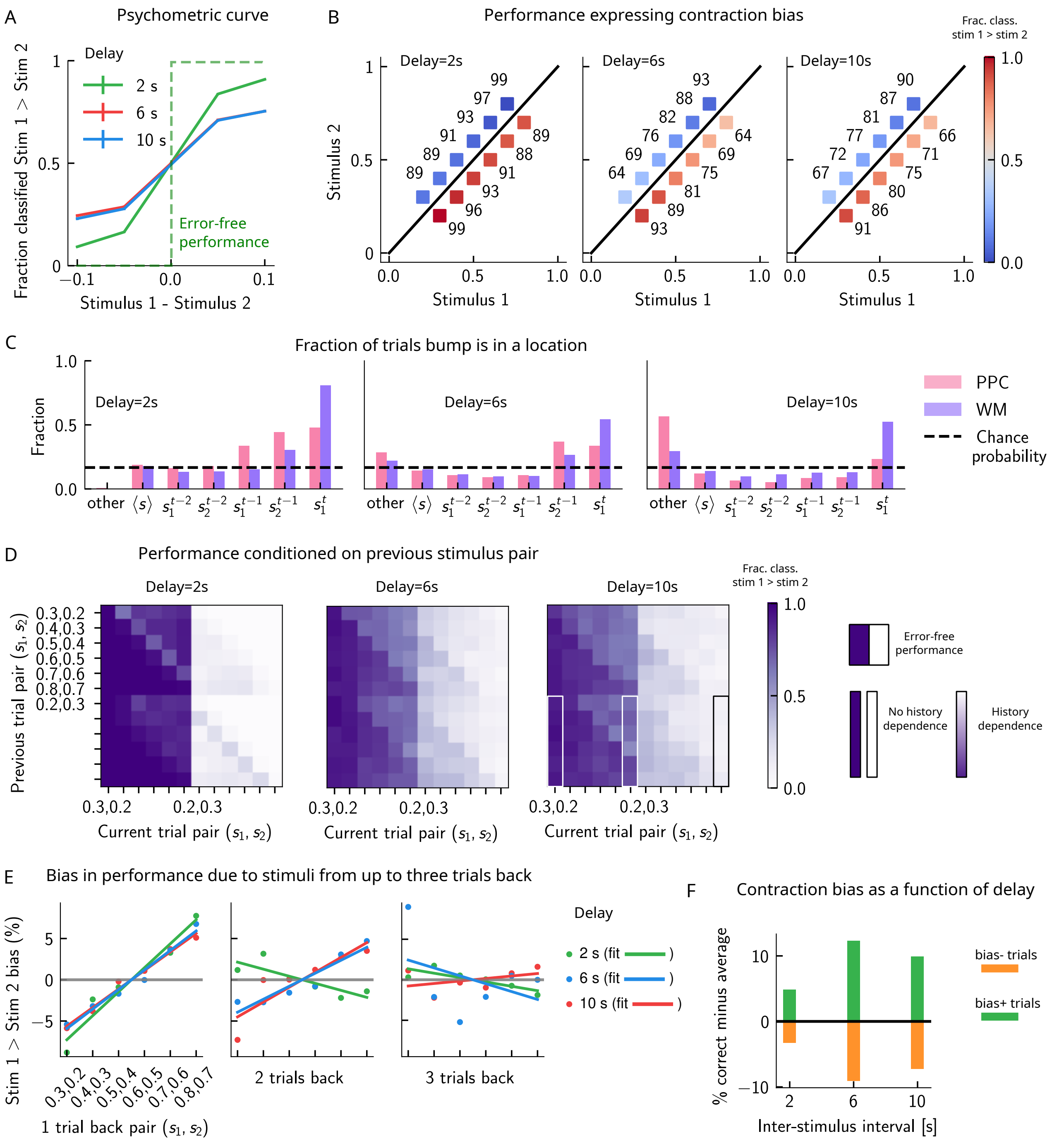

Figure S4: Model predictions for a block design. **(A)** As in (Fig. 2 A). Performance of the network model for the psychometric stimuli improves with a short delay interval and worsens as this delay is increased. **(B)** As in (Fig. 2 C). Performance is affected by contraction bias – a gradual accumulation of errors for stimuli below (above) the diagonal upon increasing (decreasing) s_1. As the delay interval increases, the contraction bias is increased which results in reduced performance across all pairs. Colorbar indicates the fraction of trials classified as $s_1 > s_2$. **(C)** As in (Fig. 3 C). The location of the bump that corresponds to the value of s_1 occupies a smaller fraction of trials, as the delay interval increases. **(D)** As in (Fig. 2 D). Performance is affected by the previous stimulus pairs (modulation along the y-axis), and becomes worse as the delay interval is increased. The colorbar corresponds to the fraction classified $s_1 > s_2$. **(E)** As in (Fig. 3 F). Bias, quantifying the (attractive) effect of the previous stimulus pairs, each color corresponding to a different delay interval. These history effects are attractive: the larger the previous trial stimulus pair, the higher the probability of classifying the first stimulus s_1 as large, and vice-versa. Middle/right panels: same as the left panel, for stimuli extending two and three trials back. **(F)** Quantifying contraction bias separately for Bias+ trials (green) and Bias- trials (orange) yields an increasing bias as the inter-stimulus interval increases.

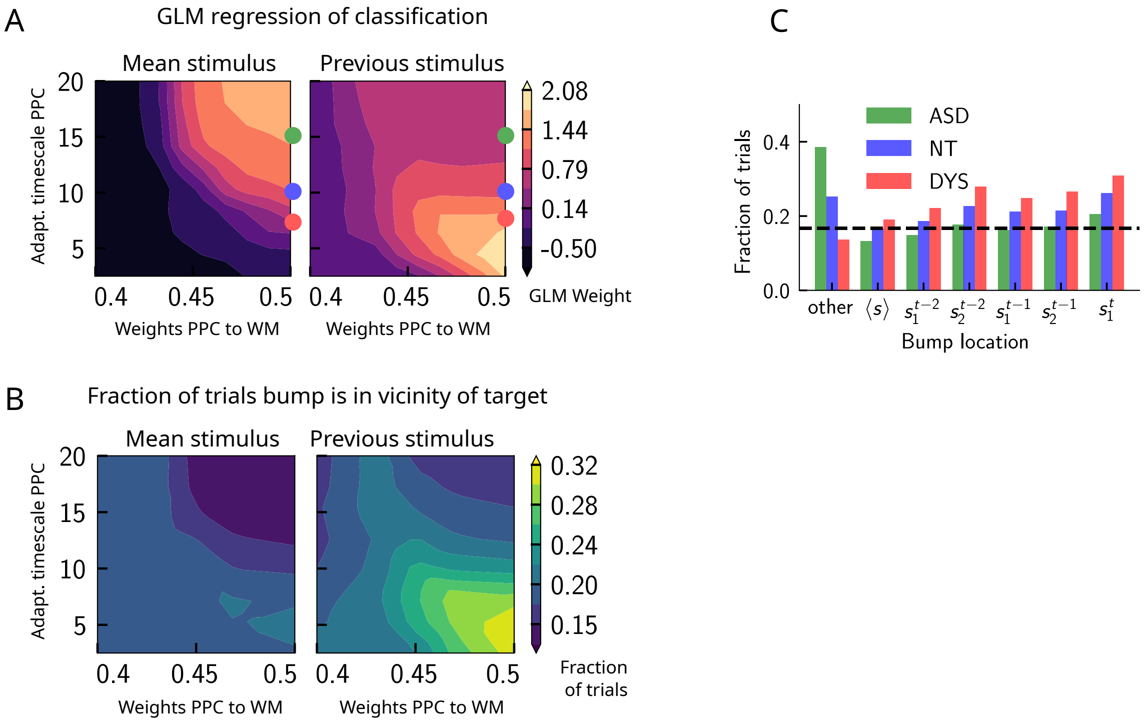

Figure S5: Apparent trade-off between short- and long-term biases, controlled by the timescale of neural adaptation. (A) Left: GLM weight associated with the regressor corresponding to the mean stimulus across trials (value indicated by colorbar), as a function of the strength of the weights from the PPC to the WM network (x-axis), and the adaptation timescale in the PPC (y-axis). Right: Same as left panel, but displaying the GLM weight associated with the regressor corresponding to the previous trial's stimulus. These two panels indicate that the adaptation timescale *seemingly* exerts a trade-off between the two biases: while decreasing it increases short-term sensory history biases, increasing it increases long-term sensory history biases. The values of the adaptation parameter marked by the three colored dots (in red, blue and green) can mimic behaviors similar to dyslexic, neurotypical, and autistic spectrum subjects (see also Fig. 7). **(B)** Left: phase diagram of the fraction of trials in which the activity bump at the end of the delay interval is in the location of the running mean stimulus as a function of the strength of the weights from the PPC to the WM network (x-axis), and the adaptation timescale in the PPC (y-axis). Right: Same as (left), but for the location of any of the two stimuli presented in the previous trial. **(C)** The fraction of trials in which the activity bump at the end of the delay interval corresponds to different locations shown in the x-axis, for three different values of the adaptation timescale parameter, corresponding to qualitatively similar to dyslexic, neurotypical, and autistic spectrum subjects, shown in colors.

4.4 Parameters

Table 1: Simulation parameters, when not explicitly mentioned. Used to produce Figs. 1, 2, 3, 6, S2, S3, S4.

Parameter	Symbol	Default value
Number of neurons	N	2000
Neuronal gain	β	5
Range of excitatory interactions [in units of stimulus space length]	d_0	0.02
Strength of inhibitory weights	J_0	0.2
Strength of excitatory weights	J_e	1
Time scale of neuronal integration in WM net [s]	τ^W	0.01
Time scale of neuronal integration in PPC [s]	τ^P	0.5
Time scale of neuronal adaptation in PPC [s]	τ_θ^P	7.5
Amplitude of adaptation current in PPC	D^P	0.3
Amplitude of external inputs	I_{ext}	1
Strength of weights from PPC to WM net	$J^{P\to W}$	0.5
Duration of stimuli [s]		0.4
Delay interval [s]		$[2, 6, 10]$
Intertrial interval [s]		5
Width of box stimulus [in units of stimulus space length]	δs	0.05

Table 2: Simulation parameters Fig. S1.

Parameter	Symbol	Default value
Number of neurons	N	1000
Neuronal gain	β	5
Range of excitatory interactions [in units of stimulus space length]	d_0	0.02
Strength of inhibitory weights	J_0	0.2
Strength of excitatory weights	J_e	1
Time scale of neuronal integration [s]	τ^P	0.01
Amplitude of external inputs	I_{ext}	1
Duration of stimuli [s]		0.4
Width of box stimulus [in units of stimulus space length]	δs	0.05

Table 3: Simulation parameters Fig. S5. Other parameters as in Tab. 1

Parameter	Symbol	Default value
Time-scale of neuronal adaptation in PPC [s]	τ_θ^P	7.5 (DYS), 10 (NT), 15 (ASD)
Amplitude of adaptation in PPC	D^P	0.2
Strength of weights from PPC to WM net	$J^{P\to W}$	0.5
Delay interval [s]		$[2, 6, 10]$
Intertrial interval [s]		1.2

Acknowledgements

We are grateful to Loreen Hertäg for helpful comments on a previous version of this manuscript, and Arash Fassihi for helpful discussions. This work was supported by BBSRC BB/N013956/1, BB/N019008/1, Wellcome Trust 200790/Z/16/Z, Simons Foundation 564408, EPSRC EP/R035806/1, Gatsby Charitable Foundation 562980 and Wellcome Trust 562763.

References

[1] Harry Levi Hollingworth. The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, 7(17):461–469, 1910.

[2] Daniel Algom. 8 memory psychophysics: An examination of its perceptual and cognitive prospects. In *Advances in psychology*, volume 92, pages 441–513. Elsevier, 1992.

[3] JE Berliner, NI Durlach, and LD Braida. Intensity perception. vii. further data on roving-level discrimination and the resolution and bias edge effects. *The Journal of the Acoustical Society of America*, 61(6):1577–1585, 1977.

[4] Åke Hellström. The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*, 97(1):35, 1985.

[5] Eustace Christopher Poulton and Simon Poulton. *Bias in quantifying judgements*. Taylor & Francis, 1989.

[6] Jerwen Jou, Gary E Leka, Dawn M Rogers, and Yolanda E Matus. Contraction bias in memorial quantifying judgment: Does it come from a stable compressed memory representation or a dynamic adaptation process? *The American journal of psychology*, pages 543–564, 2004.

[7] Paymon Ashourian and Yonatan Loewenstein. Bayesian inference underlies the contraction bias in delayed comparison tasks. *PloS one*, 6(5):e19551, 2011.

[8] Itay Lieder, Vincent Adam, Or Frenkel, Sagi Jaffe-Dax, Maneesh Sahani, and Merav Ahissar. Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nature neuroscience*, 22(2):256–264, 2019.

[9] Claudia Preuschhof, Torsten Schubert, Arno Villringer, and Hauke R Heekeren. Prior information biases stimulus representations during vibrotactile decision making. *Journal of Cognitive Neuroscience*, 22(5):875–887, 2010.

[10] Maria Olkkonen, Patrice F McCarthy, and Sarah R Allred. The central tendency bias in color perception: Effects of internal and external noise. *Journal of vision*, 14(11):5–5, 2014.

[11] Arash Fassihi, Athena Akrami, Vahid Esmaeili, and Mathew E Diamond. Tactile perception and working memory in rats and humans. *Proceedings of the National Academy of Sciences*, 111(6):2331–2336, 2014.

[12] Athena Akrami, Charles D Kopec, Mathew E Diamond, and Carlos D Brody. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature*, 554(7692):368–372, 2018.

[13] Ofri Raviv, Merav Ahissar, and Yonatan Loewenstein. How recent history affects perception: the normative approach and its heuristic approximation. *PLoS Comput Biol*, 8(10):e1002731, 2012.

[14] Jason Fischer and David Whitney. Serial dependence in visual perception. *Nature neuroscience*, 17(5):738–743, 2014.

[15] João Barbosa and Albert Compte. Build-up of serial dependence in color working memory. *Scientific reports*, 10(1):1–7, 2020.

[16] Ranulfo Romo and Emilio Salinas. Flutter discrimination: neural codes, perception, memory and decision making. *Nature Reviews Neuroscience*, 4(3):203–218, 2003.

[17] Anastasia Kiyonaga, Jason M Scimeca, Daniel P Bliss, and David Whitney. Serial dependence across perception, attention, and memory. *Trends in Cognitive Sciences*, 21(7):493–497, 2017.

[18] Guido Marco Cicchini, Kyriaki Mikellidou, and David Burr. Serial dependencies act directly on perception. *Journal of vision*, 17(14):6–6, 2017.

[19] Stefan Czoschke, Cora Fischer, Julia Beitner, Jochen Kaiser, and Christoph Bledowski. Two types of serial dependence in visual working memory. *British Journal of Psychology*, 110(2):256–267, 2019.

[20] David Alais, Garry Kong, Colin Palmer, and Colin Clifford. Eye gaze direction shows a positive serial dependency. *Journal of vision*, 18(4):11–11, 2018.

[21] Mauro Manassi, Alina Liberman, Anna Kosovicheva, Kathy Zhang, and David Whitney. Serial dependence in position occurs at the time of perception. *Psychonomic Bulletin & Review*, 25(6):2245–2253, 2018.

[22] Mauro Manassi, Alina Liberman, Wesley Chaney, and David Whitney. The perceived stability of scenes: serial dependence in ensemble representations. *Scientific reports*, 7(1):1–9, 2017.

[23] Marta Suárez-Pinilla, Anil K Seth, and Warrick Roseboom. Serial dependence in the perception of visual

variance. *Journal of Vision*, 18(7):4–4, 2018.

[24] Charalampos Papadimitriou, Afreen Ferdoash, and Lawrence H Snyder. Ghosts in the machine: memory interference from the previous trial. *Journal of neurophysiology*, 113(2):567–577, 2015.

[25] Yonatan Loewenstein, Ofri Raviv, and Merav Ahissar. Dissecting the roles of supervised and unsupervised learning in perceptual discrimination judgments. *Journal of Neuroscience*, 41(4):757–765, 2021.

[26] H Sebastian Seung. Continuous attractors and oculomotor control. *Neural Networks*, 11(7-8):1253–1258, 1998.

[27] Xiao-Jing Wang. Synaptic reverberation underlying mnemonic persistent activity. *Trends in neurosciences*, 24(8):455–463, 2001.

[28] John D Murray, Alberto Bernacchia, David J Freedman, Ranulfo Romo, Jonathan D Wallis, Xinying Cai, Camillo Padoa-Schioppa, Tatiana Pasternak, Hyojung Seo, Daeyeol Lee, et al. A hierarchy of intrinsic timescales across primate cortex. *Nature neuroscience*, 17(12):1661–1663, 2014.

[29] Adrian Hernandez, Emilio Salinas, Rafael Garcıa, and Ranulfo Romo. Discrimination in the sense of flutter: new psychophysical measurements in monkeys. *Journal of Neuroscience*, 17(16):6391–6400, 1997.

[30] Robert J Sinclair and Harold Burton. Discrimination of vibrotactile frequencies in a delayed pair comparison task. *Perception & psychophysics*, 58(5):680–692, 1996.

[31] Arash Fassihi, Athena Akrami, Francesca Pulecchi, Vinzenz Schönfelder, and Mathew E Diamond. Transformation of perception from sensory to motor cortex. *Current Biology*, 27(11):1585–1596, 2017.

[32] David Hansel and Haim Sompolinsky. 13 modeling feature selectivity in local cortical circuits, 1998.

[33] Jose M Esnaola-Acebes, Alex Roxin, and Klaus Wimmer. Bump attractor dynamics underlying stimulus integration in perceptual estimation tasks. *BioRxiv*, 2021.

[34] Sagi Jaffe-Dax, Eva Kimel, and Merav Ahissar. Shorter cortical adaptation in dyslexia is broadly distributed in the superior temporal lobe and includes the primary auditory cortex. *ELife*, 7:e30018, 2018.

[35] Sagi Jaffe-Dax, Or Frenkel, and Merav Ahissar. Dyslexics' faster decay of implicit memory for sounds and words is manifested in their shorter neural adaptation. *Elife*, 6:e20557, 2017.

[36] Ke Tong and Chad Dubé. A tale of two literatures: A fidelity-based integration account of central tendency bias and serial dependency. *Computational Brain & Behavior*, pages 1–21, 2022.

[37] Muhsin Karim, Justin A Harris, Angela Langdon, and Michael Breakspear. The influence of prior experience and expected timing on vibrotactile discrimination. *Frontiers in neuroscience*, 7:255, 2013.

[38] Stephen M Kerst and James H Howard. Memory psychophysics for visual area and length. *Memory & Cognition*, 6(3):327–335, 1978.

[39] Dobromir Rahnev and Rachel N Denison. Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, 41, 2018.

[40] I Hachen, S Reinartz, R Brasselet, A Stroligo, and ME Diamond. Dynamics of history-dependent perceptual judgment. *Nature communications*, 12(1):1–15, 2021.

[41] Amadeus Maes, Mauricio Barahona, and Claudia Clopath. Long-and short-term history effects in a spiking network model of statistical learning. *bioRxiv*, 2021.

[42] Emilio Salinas. Prior and prejudice. *Nature neuroscience*, 14(8):943–945, 2011.

[43] Lux Li, Arielle Chan, Shah M Iqbal, and Daniel Goldreich. An adaptation-induced repulsion illusion in tactile spatial perception. *Frontiers in human neuroscience*, 11:331, 2017.

[44] Matthias Fritsche, Pim Mostert, and Floris P de Lange. Opposite effects of recent history on perception and decision. *Current Biology*, 27(4):590–595, 2017.

[45] Matthias Fritsche, Samuel G Solomon, and Floris P de Lange. Brief stimuli cast a long-term trace in visual cortex. *BioRxiv*, 2021.

[46] Valentina Daelli, Nicola J van Rijsbergen, and Alessandro Treves. How recent experience affects the perception of ambiguous objects. *Brain research*, 1322:81–91, 2010.

[47] Matthias Fritsche, Eelke Spaak, and Floris P De Lange. A bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *Elife*, 9:e55389, 2020.

[48] Marco Boi, Haluk Öğmen, and Michael H Herzog. Motion and tilt aftereffects occur largely in retinal, not in

object, coordinates in the ternus–pikler display. *Journal of Vision*, 11(3):7–7, 2011.

[49] Tomas Knapen, Martin Rolfs, Mark Wexler, and Patrick Cavanagh. The reference frame of the tilt aftereffect. *Journal of Vision*, 10(1):8–8, 2010.

[50] Sebastiaan Mathôt and Jan Theeuwes. A reinvestigation of the reference frame of the tilt-adaptation aftereffect. *Scientific reports*, 3(1):1–7, 2013.

[51] Michele Fornaciai and Joonkoo Park. Spontaneous repulsive adaptation in the absence of attractive serial dependence. *Journal of vision*, 19(5):21–21, 2019.

[52] Michael A Pitts, William J Gavin, and Janice L Nerger. Early top-down influences on bistable perception revealed by event-related potentials. *Brain and cognition*, 67(1):11–24, 2008.

[53] Gerald M Long, Thomas C Toppino, and Gregory W Mondin. Prime time: Fatigue and set effects in the perception of reversible figures. *Perception & psychophysics*, 52(6):609–616, 1992.

[54] Gerald M Long and Cindy J Moran. How to keep a reversible figure from reversing: Teasing out top—down and bottom—up processes. *Perception*, 36(3):431–445, 2007.

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

[56] Joaquin M Fuster and Garrett E Alexander. Neuron activity related to short-term memory. *Science*, 173(3997):652–654, 1971.

[57] Yasushi Miyashita and Han Soo Chang. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331(6151):68–70, 1988.

[58] Shintaro Funahashi, Charles J Bruce, and Patricia S Goldman-Rakic. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of neurophysiology*, 61(2):331–349, 1989.

[59] Shintaro Funahashi, Charles J Bruce, and Patricia S Goldman-Rakic. Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. *Journal of neurophysiology*, 63(4):814–831, 1990.

[60] Ranulfo Romo, Carlos D Brody, Adrián Hernández, and Luis Lemus. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, 399(6735):470–473, 1999.

[61] Emilio Salinas, Adrian Hernandez, Antonio Zainos, and Ranulfo Romo. Periodicity and firing rate as candidate neural codes for the frequency of vibrotactile stimuli. *Journal of neuroscience*, 20(14):5503–5515, 2000.

[62] Xiaoxing Zhang, Wenjun Yan, Wenliang Wang, Hongmei Fan, Ruiqing Hou, Yulei Chen, Zhaoqin Chen, Chaofan Ge, Shumin Duan, Albert Compte, et al. Active information maintenance in working memory by a sensory cortex. *Elife*, 8:e43191, 2019.

[63] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[64] Daniel J Amit and Daniel J Amit. *Modeling brain function: The world of attractor neural networks*. Cambridge university press, 1992.

[65] Francesco P Battaglia and Alessandro Treves. Stable and rapid recurrent processing in realistic autoassociative memories. *Neural Computation*, 10(2):431–450, 1998.

[66] Omri Barak, David Sussillo, Ranulfo Romo, Misha Tsodyks, and LF Abbott. From fixed points to chaos: three models of delayed discrimination. *Progress in neurobiology*, 103:214–222, 2013.

[67] Gianluigi Mongillo, Omri Barak, and Misha Tsodyks. Synaptic theory of working memory. *Science*, 319(5869):1543–1546, 2008.

[68] Omri Barak and Misha Tsodyks. Persistent activity in neural networks with dynamic synapses. *PLoS Comput Biol*, 3(2):e35, 2007.

[69] Joao Barbosa, Heike Stein, Rebecca L Martinez, Adrià Galan-Gadea, Sihai Li, Josep Dalmau, Kirsten Adam, Josep Valls-Solé, Christos Constantinidis, and Albert Compte. Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. *Nature neuroscience*, 23(8):1016–1024, 2020.

[70] Henry WP Dalgleish, Lloyd E Russell, Adam M Packer, Arnd Roth, Oliver M Gauld, Francesca Greenstreet, Emmett J Thompson, and Michael Häusser. How many neurons are sufficient for perception of cortical activity? *Elife*, 9:e58889, 2020.