PNAS RESEARCH ARTICLE | NEUROSCIENCE

Learning predictive signals within a local recurrent circuit

Toshitake Asabuki[a,b,c,1], Colleen J. Gillon[a], and Claudia Clopath[a,1]

Edited by Karl J. Friston, University College London, London, United Kingdom; received July 22, 2024; accepted May 3, 2025 by Editorial Board Member Peter L. Strick

The predictive coding hypothesis proposes that top–down predictions are compared with incoming bottom–up sensory information, with prediction errors signaling the discrepancies between these inputs. While this hypothesis explains the presence of prediction errors, recent experimental studies suggest that prediction error signals can emerge within a local circuit, that is, from bottom–up sensory input alone. In this paper, we test whether local circuits alone can generate predictive signals by training a recurrent spiking network using local plasticity rules. Our network model replicates experimentally observed features of prediction errors, such as biphasic neural activity patterns and context dependency. Our findings shed light on how synaptic plasticity can shape prediction errors and enable the acquisition and updating of an internal model of sensory input within a recurrent neural network.

synaptic plasticity | recurrent spiking network | predictive coding | prediction error signal

Significance

The brain operates as a predictive system, using an internal model to predict upcoming sensory inputs under uncertain conditions. Although prediction error signals, which reflect the discrepancy between predicted and actual sensory events, are a crucial component of predictive processing, the neural mechanisms underlying these signals remain elusive. Using a computational model, we investigate synaptic plasticity rules that learn prediction error signals in a local recurrent circuit. We find that recurrent networks trained with the proposed plasticity rules explain many features of prediction errors observed in experimental studies. Our study clarifies the basic requirements for learning with prediction error signals in recurrent networks, a crucial step toward understanding how the brain constructs its internal model of the environment.

The brain is thought to learn an internal model of the environment to predict upcoming sensory inputs (1). In support of this hypothesis, a wide variety of experiments have reported mismatch responses in the brain (2–4). These responses are typically elicited by presenting subjects with a series of familiar or consistent stimuli, and then introducing an unexpected stimulus. Mismatch responses then emerge as the difference between the neural activity evoked by the expected stimulus and the unexpected stimulus. For example, using electroencephalography (EEG), unexpected stimuli have been shown to elicit a mismatch negativity (MMN) response in humans, typically seen as a biphasic response in which a negative deflection is followed by a positive one (2, 5–9). Early studies of this mismatch response were primarily conducted using auditory oddball paradigms (2, 5). However, many studies have since shown MMN-like responses emerging in a variety of sensory tasks and brain regions, including visual (9–13) and auditory (14, 15) areas, as well as cognitive processing regions like prefrontal cortex (16, 17). Although there may be some differences between these signals, together this suggests that mismatch responses may constitute a general mechanism for automatically detecting deviations from the brain's internal model.

A plausible explanation for these mismatch signals in the brain is provided by the predictive coding hypothesis (2). In predictive coding, at each level of sensory processing, top–down predictions from the brain's internal model of the world are used to cancel out incoming sensory information. Only the discrepancies between the predicted and actual sensory information are communicated to higher levels of the cortical hierarchy. These discrepancies are called prediction errors and they are thought to be critical for the brain to improve its internal model, and thus the predictions passed down through the sensory processing hierarchy (18–20). Several models derived from the predictive coding hypothesis have been implemented using biologically plausible models of neurons to explain the mismatch signals observed in the brain. For example, Wacongne et al. (3) proposed a spiking neuron model of predictive coding to account for the MMN in an oddball paradigm (3). Relatedly, a study by Lieder et al. (21) based on dynamic causal models provided a potential link between the MMN and Bayesian filtering (21). In these models, mismatch responses are explained to be prediction errors, computed at different levels of the sensory processing hierarchy.

Although these types of generative predictive coding models provide a broad explanation of the emergence of mismatch signals, there are two important aspects that they do not capture. First, there is considerable evidence that the brain develops internal predictions through experience, and thus that learning and prediction errors occur alongside one another (5, 7, 22–24). While several studies have attempted to demonstrate how a set of synaptic plasticity rules can account for both prediction and error signaling within a network (25–28), it is still unclear whether the prediction error signals that emerge at the population level after learning can account for those observed in experimental data. Second, several studies have shown that mismatch responses can occur automatically in

Author affiliations: [a]Department of Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom; [b]RIKEN Center for Brain Science, Wako, Saitama 351-0198, Japan; and [c]RIKEN Pioneering Research Institute, Wako, Saitama 351-0198, Japan

Author contributions: T.A., C.J.G., and C.C. designed research; T.A. performed research; T.A. analyzed data; and T.A., C.J.G., and C.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. K.J.F. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: toshitake.asabuki@riken.jp or c.clopath@imperial.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2414674122/-/DCSupplemental.

Published July 1, 2025.

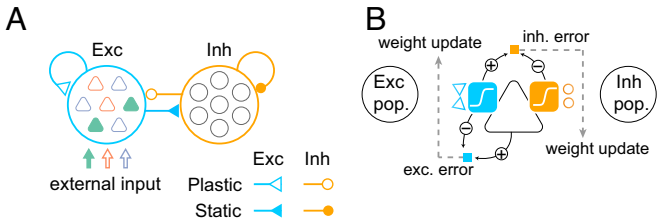

Fig. 1. Model. (*A*) A network model with distinct excitatory and inhibitory populations. Only excitatory populations are driven by external inputs. Only synapses that project to excitatory neurons are set to be plastic. (*B*) A schematic of the plasticity rules proposed in ref. 29. Excitatory (blue) and inhibitory (orange) synapses projecting to an excitatory neuron (triangle) obey different plasticity rules. For excitatory synapses, errors between internally driven excitation (blue sigmoid) and the output of the cell provide feedback to the synapses (dashed arrow) and modulate plasticity (blue square; exc. error). All excitatory connections seek to minimize these errors. For inhibitory synapses, the error between internally driven excitation (blue sigmoid) and inhibition (orange sigmoid) must be minimized to maintain excitation–inhibition balance (orange square; inh. error). Note that although excitatory and inhibitory potentials were passed through distinct nonlinearities in the plasticity rule, actual membrane potentials were calculated using the same sigmoid function (Eqs. **3** and **4**).

contexts during which there appears to be minimal top–down input, like when subjects are inattentive or even sleeping (2). However, most existing models focus on the emergence of mismatch signals in hierarchical circuits, and do not explain how such signals can emerge within a local circuit, from bottom–up sensory input alone.

In this paper, we addressed these open questions by simulating a recurrent spiking network using local plasticity rules we recently proposed (29). The network model consists of a population of excitatory and inhibitory spiking neuron models. The synapses onto excitatory neurons undergo synaptic plasticity, allowing them to develop connectivity patterns that predict the network activity evoked by upcoming sensory events. Simultaneously, inhibitory synapses undergo additional plasticity to maintain the excitatory-inhibitory balance. We found that recurrent networks trained with these plasticity rules replicated many features of prediction errors that have been observed in experimental studies (2, 5–9). For example, the prediction errors displayed a similar biphasic pattern to the MMN waveform, and were context-dependent (30, 31). Overall, our study provides insights into the mechanisms by which synaptic plasticity shapes prediction errors, and the acquisition and updating of an internal model of sensory input within a recurrent neural network.

Results

To test whether predictive signals can be computed in a local circuit, we simulated a recurrent spiking network consisting of excitatory (E) and inhibitory (I) model neurons (Fig. 1*A*). Only excitatory neurons were driven by external stimuli. Neurons were assumed to generate spikes stochastically, mimicking the noisy fluctuations of membrane potentials. We presented a number of stimuli to the network, each of which increased the firing rate of a nonoverlapping subset of excitatory neurons (see *SI Appendix*, Fig. S3 for overlapping assemblies). All feedforward connections were held fixed.

We investigated how prediction errors are formed through sensory experiences by using synaptic plasticity rules that we proposed previously (29). We designed the model such that excitatory and inhibitory synapses undergo distinct plasticity rules. Briefly, excitatory synapses that contributed to predicting neural activity were strengthened (32–35) (Fig. 1*B*, blue square; Eqs. **8** and **9** in *Methods*), while the inhibitory synapses were modified to maintain the excitation–inhibition balance (EI balance) by learning to predict the recurrent excitatory potential (Fig. 1*B*, orange square; Eqs. **10** and **11** in *Methods*). In the following, we describe the behavior of each plasticity rule more specifically.

The change in excitatory synaptic strength between presynaptic neuron j and postsynaptic neuron i is proportional to the error between the instantaneous excitatory firing rate of the postsynaptic neuron f_i^E and the total internally driven excitatory inputs filtered by the sigmoidal function y_i^E as:

$$\Delta W_{ij}^{EE} = \epsilon \left[f_i^E - y_i^E \right] \bullet x_j^E,$$

where x_j^E is a postsynaptic potential evoked by excitatory neuron j (Eqs. **6** and **7**). Under this rule, in the case of a positive error, the synapse undergoes long-term potentiation (LTP) and in the case of a negative error, the synapse undergoes long-term depression (LTD) (*SI Appendix*, Fig. S1*A*).

Similarly, for the inhibitory plasticity, the change in synaptic strength between presynaptic neuron j and postsynaptic neuron i is proportional to the error between the sigmoidal of the total excitatory y_i^E and inhibitory y_i^I inputs as:

$$\Delta W_{ij}^{EI} = \epsilon \left[y_i^E - y_i^I \right] x_j^I,$$

where x_j^I is the postsynaptic potential evoked by inhibitory neuron j (Eqs. **6** and **7**). As under the excitatory plasticity rule, inhibitory synapses were updated when the presynaptic activity and the error term coincided, with a positive error leading to LTP while a negative one led to LTD (*SI Appendix*, Fig. S1*B*).

Emergence of a Biphasic Prediction Error in a Local Recurrent Network. Numerous experimental studies using EEG and magneto-encephalography have shown that a mismatch signal arises in the auditory cortex when a rare "unexpected" auditory stimulus occurs among a sequence of consistently repeated "expected" stimuli (2, 36, 37). This mismatch signal is measured by subtracting the response to the expected event from the response to the unexpected one. Typically, this response difference (also known as a difference wave) comprises both a negative (MMN) and a positive component (2, 5–9). Intriguingly, further experimental studies have found that the inferotemporal cortex shows similar biphasic mismatch responses when a violation of transitional rules imposed during learning occurs (9). Despite the consistency of these observations over various tasks, the plasticity mechanism that generates biphasic mismatch response is still unclear.

We first asked whether the proposed model could account for this biphasic mismatch response when a transition is violated in a learned sequence. To this end, a sequence with the deterministic transition "ABC" was presented to the network during a learning phase (Fig. 2 *A*, *Top*). The excitatory synapses within each assembly (i.e., group of neurons targeted by the same stimulus, e.g., "A") increased in strength through learning, indicating the formation of cell assemblies for all stimuli (Fig. 2*B*, diagonal blocks). Second, between-assembly connections for assembly A to B and B to C were strengthened, indicating that the model learned the transition probabilities between stimulus patterns as we have shown previously (Fig. 2*B*, blue squares) (29).

We then investigated whether the network which learned stereotypical sequences showed a prediction error signal when an unexpected sequence was presented. To this end, we measured the entire network response over an expected sequence (ABC) and a sequence with an unexpected transition (ABA) (Fig. 2 *A*, *Bottom*). In this analysis, all synaptic weights were fixed so that

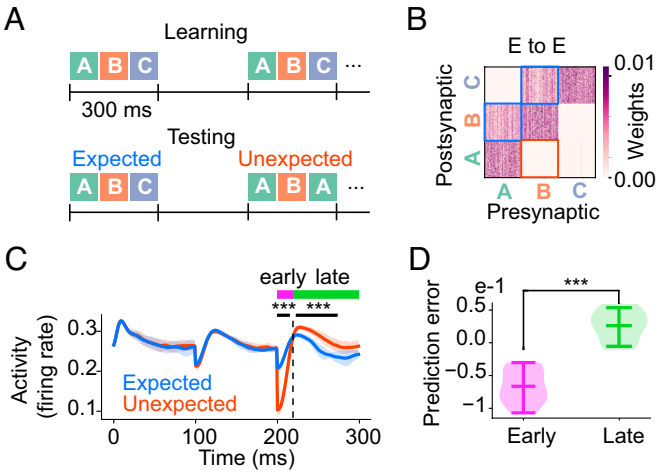

Fig. 2. Biphasic prediction error learned through plasticity. (*A*, *Top*) During learning, the sequence ABC was repeatedly presented to the network. We included 300 ms-long gaps between each sequence. (*A*, *Bottom*) After learning, all synapses were fixed and both the expected sequence ABC and the unexpected sequence "ABA" were presented alternately. (*B*) Learned excitatory synapses are shown. Synapses were strengthened within each assembly (diagonal component of the matrix) and between assemblies that had transitions in the expected sequence (blue squares). Red squares show synapses between assemblies in the unexpected sequence. (*C*) Mean firing rates of the whole network during the expected (blue) and unexpected (red) sequences are shown. Period during which the last elements of sequences were presented was divided into early and late phases. Shaded areas represent s.d. over 10 trials. Black horizontal lines show periods during which the two responses showed a significant difference. (*D*) Mean prediction errors during early and late phases over 10 independent simulations are shown. Here, prediction error was defined as difference between responses to unexpected and expected sequences (unexpected − expected). In *C* and *D*, *P*-values were calculated using a two-sided Welch's *t* test (***$P < 0.001$).

we could monitor the pure dynamics of the network. Note that the transition from pattern B to A in an unexpected sequence violated the transition rule established during learning, and hence recurrent excitation connections from B to A had not been enhanced (Fig. 2*B*, red square). In both the expected and unexpected case, the network firing rates immediately after a transition showed an abrupt drop (early phase) followed by a slower rise (late phase) (Fig. 2*C*, around vertical dashed line). However, we found a significant difference when the transition violated the expected sequence, in the unexpected case: the response amplitudes were much stronger in both the negative and positive phases, making the resulting error signals biphasic (Fig. 2*D*; membrane potentials shown in *SI Appendix*, Fig. S2), consistent with results reported in the EEG literature (2, 3, 5, 6, 8, 9). We have confirmed that the biphasic error signal still exists even when the assemblies have shared memberships (*SI Appendix*, Fig. S3 *A* and *B*), as the nonoverlapping part of the assemblies learn the sequence structure well (*SI Appendix*, Fig. S3*C*). Furthermore, the model shows biphasic error responses even if the expected sequences allow for multiple stochastic transitions (*SI Appendix*, Fig. S4).

We also asked whether the proposed plasticity rule learns to generate biphasic prediction error signals in a simple oddball paradigm, as previous predictive coding algorithms have shown (3). In an oddball paradigm, two stimuli are presented: a frequent stimulus and a rare, "oddball" stimulus. We implemented this paradigm by presenting our model with sequences either comprising only the frequent stimulus (AAAAA) or ending with a rare stimulus (AAAAB) (*SI Appendix*, Fig. S5*A*). Since the two different sequences were presented with equal probability, stimulus A was the frequent stimulus, and stimulus B was the rare or oddball stimulus. We will call these stimuli the expected and unexpected stimulus, respectively. We found that, after the network had been trained with the sequences, it showed a biphasic prediction error response to the unexpected stimulus compared to the expected stimulus (*SI Appendix*, Fig. S5 *B* and *C*).

In summary, these results show that our network model learns prediction error responses when presented with a stimulus sequence transition violation. In particular, the model shows a biphasic prediction error response, comprising a negative and a positive component, as found in neurophysiological experiments.

Network Mechanism of Biphasic Prediction Errors. Returning to the ABC sequences, we next asked what network mechanism underlies the biphasic mismatch signal observed (Fig. 2). As the network dynamics were determined by recurrent connections between and within assemblies, we analyzed the dynamics of the excitatory and inhibitory recurrent currents. Here, we limited our analysis to the period during which the last stimulus of each sequence was presented, as the prediction error occurs only within this period. Specifically, we analyzed the currents in assembly C for the expected case and in assembly A for the unexpected case (*SI Appendix*, Fig. S6). We first explain the mechanism underlying the negative component of the prediction error, which occurs during the early phase. During the early phase in the expected sequence, excitatory and inhibitory currents showed similar levels, indicating that the model approximately maintained EI balance (Fig. 3*A*, expected). In contrast, in the unexpected case, these currents showed a significant difference (Fig. 3*A*, unexpected; Fig. 3*B*), breaking the EI balance. As the inhibitory current in the unexpected case was dominant over the excitatory current (Fig. 3*B*), a negative component appeared in the prediction error in the early phase. Note that this break in EI balance was triggered by a drastic decrease in excitatory currents in the unexpected compared to the expected case (Fig. 3*A*, cyan). Although the inhibitory currents showed a similar pattern (Fig. 3*A*, orange), the difference was much smaller than for the excitatory currents (Fig. 3*C*). The significant difference in excitatory currents is likely explained by the fact that the recurrent connections from assembly B to A were not strengthened during learning, as we have already seen (Fig. 2*B*).

The positive error component during the late phase is more surprising. We analyzed recurrent currents for both excitatory and inhibitory neurons during the late phase, as we had done for the early phase. As in the early phase, the EI balance was maintained in the expected case (Fig. 3*D*, expected) and was broken in the unexpected case (Fig. 3*D*, unexpected; Fig. 3*E*). Notably, we found that, in contrast to the early phase, the excitatory currents were disproportionately stronger than the inhibitory currents in the unexpected case, generating a positive prediction error during the late phase. Thus, the EI balance was broken in the opposite direction due to a significant decrease in inhibitory currents in the unexpected case compared to the expected one (Fig. 3*F*).

To further understand the origin of the negative prediction error measured in the early phase, we also analyzed the total incoming recurrent currents from each assembly. The negative prediction error during the early phase is the result of the suppression of excitatory synaptic input in the unexpected case (Fig. 3*C*). By measuring the total excitatory recurrent inputs provided by each assembly (Fig. 3*G*), we found that assembly C in the expected sequence was strongly activated by assemblies B and C, whereas assembly A in the unexpected sequence was solely activated by itself. This explains the observed decrease in excitatory currents in the unexpected case. It is due to the fact that strong excitatory connections would have formed from assembly B to C during the

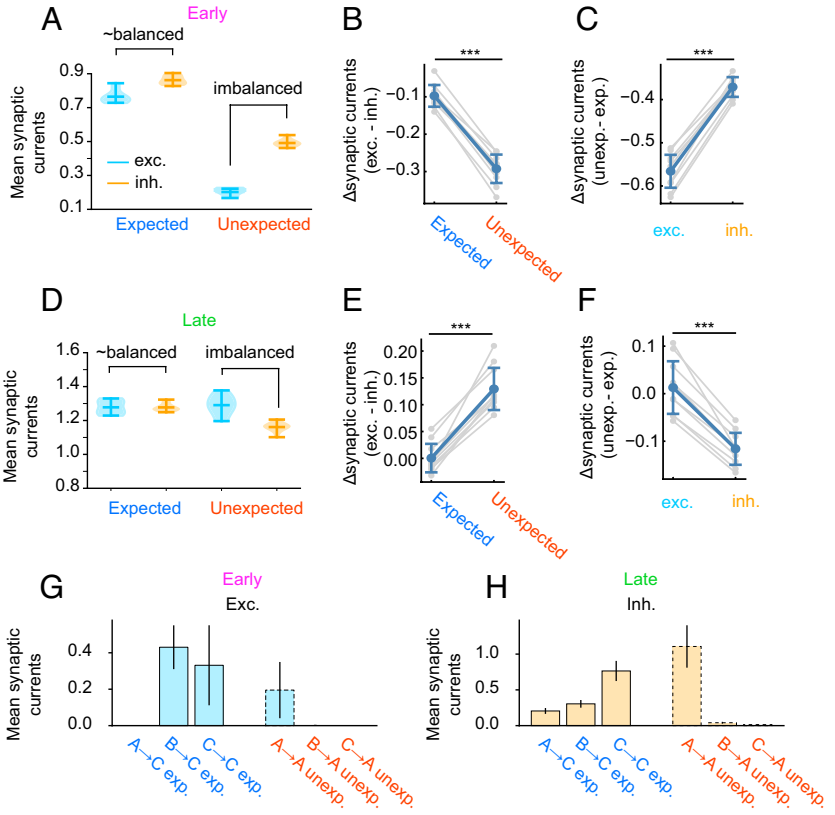

Fig. 3. Analysis of recurrent currents. (*A*) Averaged excitatory (cyan) and inhibitory (orange) currents in the early phase for expected and unexpected cases are shown. The symbol "~" indicates that the absolute value of the current change is less than 0.1 and an approximate EI balance is maintained in the expected case. (*B*) Differences between excitatory and inhibitory currents in the early phase for the expected and unexpected cases are shown. (*C*) Differences in excitatory (cyan) or inhibitory (orange) synaptic currents between the unexpected and expected cases (unexpected − expected) in the early phase are shown. (*D*) Same as *A*, but for the late phase. (*E*) Same as *B*, but for the late phase. (*F*) Same as *C*, but for the late phase. (*G*) Mean excitatory synaptic currents projecting onto neurons in assembly C in the expected case or onto assembly A in the unexpected case are shown. (*H*) Mean inhibitory synaptic currents in the late phase, projecting onto neurons in assembly C in the expected case or onto A in the unexpected case are shown. In *B*, *C*, *E*, and *F*, *P*-values were calculated using a two-sided Welch's *t* test (***$P < 0.001$). Data points for each case were generated by 10 independent simulations.

learning phase, but not from assembly B to A. The strong self-inhibiting currents in assembly A provide another mechanism: the negative error response of assembly A during the early phase leads to self-disinhibition of the assembly, which in turn enhances the positive error response.

We then repeated this analysis for the late phase. As this positive component was the result of suppressed inhibitory input in the unexpected case (Fig. 3*F*), we calculated total inhibitory recurrent inputs over different pairs of assemblies (Fig. 3*H*). We found that assembly C in the expected case was inhibited by all assemblies, whereas assembly A in the unexpected sequence was inhibited only by itself.

Altogether, these results suggest that negative and positive prediction errors in the early and late phases, respectively, result from distinct mechanisms. They show further that both negative and positive prediction errors can be explained by a disruption in the EI balance in the unexpected case, but that the underlying mechanism is different for the two phases: in the early phase, the EI balance is broken due to significantly reduced excitatory currents, whereas in the late phase, the break is due to reduced inhibitory currents.

Interplay of Excitatory and Inhibitory Plasticity in Shaping Prediction Error Signals. To validate whether the proposed plasticity rules are necessary to generate such biphasic prediction error signals in our model, we then compared network dynamics over multiple conditions: a) fixed recurrent connections, b) no recurrent connections, c) fixed excitatory, but plastic inhibitory connections, and d) fixed inhibitory, but plastic excitatory connections. In both the fixed recurrent connections case (i.e., case a) and the no recurrent connections case (i.e., case b), unexpected transitions did not lead to significant differences in activity compared to the expected case (*SI Appendix*, Fig. S7 *A* and *B*). In the fixed excitatory connections case (i.e., case c), we found that the unexpected transition led to suppressed activity compared to the expected transition, resulting only in a negative prediction error (*SI Appendix*, Fig. S7*C*). Finally, in the fixed inhibitory connections case (i.e., case d), we found that the model did still show a biphasic error response (*SI Appendix*, Fig. S7*D*). However, there was an important difference between this case and the original experiment case (i.e., Fig. 2*C*): in case d, the positive prediction error was not as sustained as in the original experiment case. This is because the original positive prediction error signal resulted from the maturation of the inhibitory synapses in response to training with the expected sequence, as seen in Fig. 3*H*.

Learning of Expectation-Dependent Prediction Error Signals. The results described above show that the model displays a transitional surprise response if a predicted stimulus in a sequence is replaced by another stimulus, creating an unexpected transition. If the network does indeed come to encode expected sequence order through learning, responses to a same stimulus should be influenced by how predictable it was given the stimuli that preceded it.

In this vein, a recent experimental study found evidence that the primary visual cortex (V1) responds differently to a stimulus based on whether the preceding stimulus in the sequence was

expected or unexpected (31). In their experiment, extracellular neuronal recordings were acquired in awake head-fixed mice viewing sequences of visual stimuli "ABCD," where each stimulus in the sequence had a set orientation. Mice were randomly assigned to four test days (i.e., days 1 to 4), such that each group experienced a different total number of learning days (Fig. 4*A*). After experiencing the test stimuli once, mice were removed from the experiment. To quantify to what extent prediction certainty influences neural responses, two types of sequences (i.e., "AB̃CD" and "EB̃CD") were used as test stimulus sequences in the experiment. Here, "E" was an unexpected stimulus which was not present in the learned sequence "ABCD". It should be noted that the "B̃" notation reflects the fact that the features of the "B" shown at test time varied in their orientation by a few degrees compared to the trained "B". To summarize the experimental results, when comparing V1 neural activity in mice tested on days 1 and 4 of learning, the late responses to stimulus "B̃" in the sequences starting with "AB̃" were significantly suppressed, whereas the late responses to stimulus "B̃" in the sequences starting with "EB̃" only showed a trend of decreasing over days. The early phase responses showed no change (SI Appendix, figure S6 in ref. 31.

We sought to test whether our model is consistent with these experimental results. In our simulation, we trained the network on an "ABCD" sequence, the duration of each stimulus was 150 ms. When comparing our network to the mouse brain, it is important to consider that area V1 in adult mice is not a fully random network at the start of these experiments. It has been shaped by extensive visual experience, which, although it does not include the exact "E" stimulus, would include stimuli with overlapping features. To take this into account in our model, we also trained our network with a single stimulus sequence "E" so that at the test time, the network would have some previous experience of all the different stimuli being studied (Fig. 4 *B*, *Top*). Stimulus sequences were presented one after another with 600 ms gaps in between. As learning progressed, recurrent connections formed five assemblies, with four of them (i.e., A, B, C, and D) being connected by unidirectional projections (as in Fig. 2 and *SI Appendix*, Fig. S8). We then asked how the maturation of synaptic strength during learning shapes the prediction errors obtained by comparing the response to an expected sequence (i.e., "AB̃CD"), and an unexpected sequence ("EB̃CD"), as in the experimental setting described above (Fig. 4 *B*, *Bottom*). To reproduce the slight noise added to stimulus "B" in the experiment in creating the "B̃" stimulus, we weakly activated three out of five assemblies, while

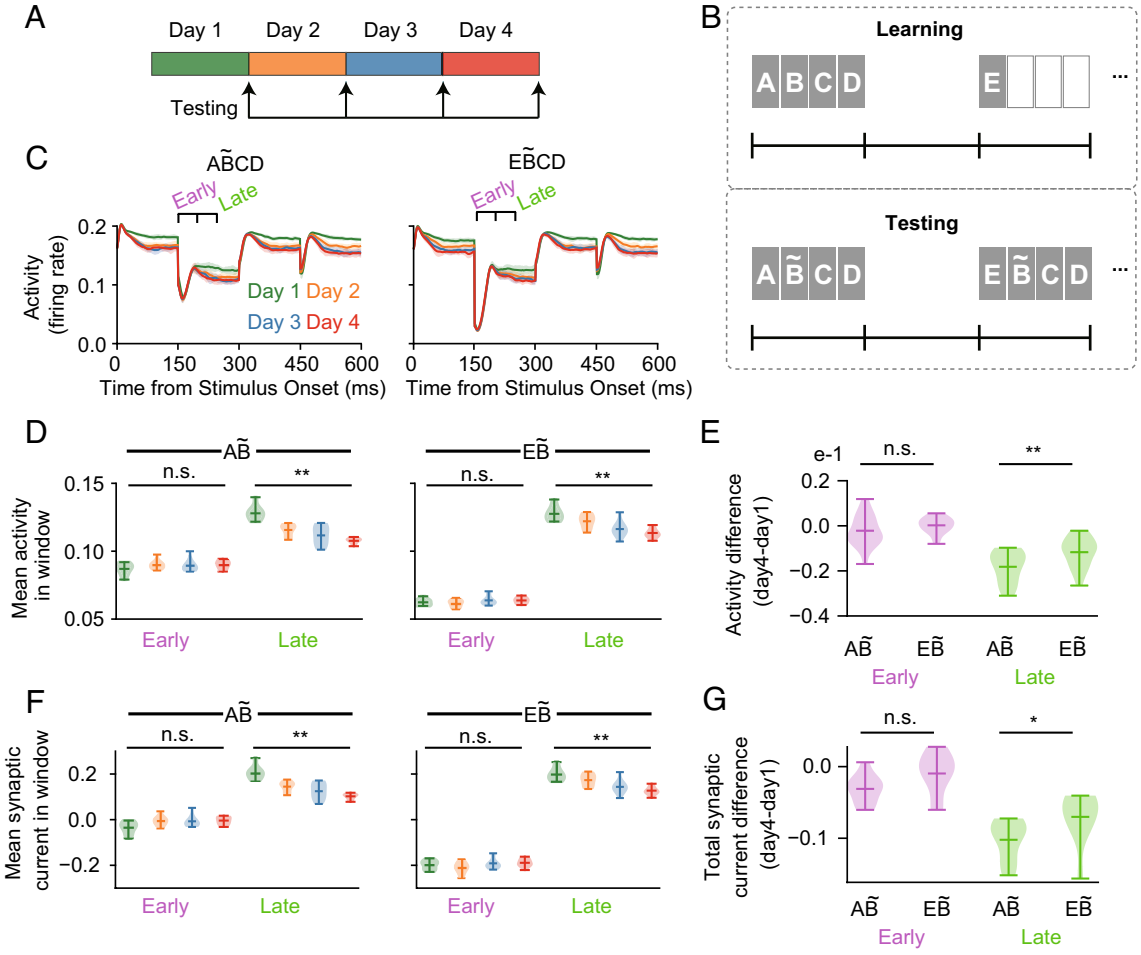

Fig. 4. Learning of expectation-dependent prediction error signals. (*A*) The learning phase was divided into four stages. At the end of each stage, all synapses were fixed and network responses were tested for expected and unexpected sequences. (*B*, *Top*) During learning, sequence ABCD and an isolated stimulus E were presented alternately. (*B*, *Bottom*) During the testing phase, two sequences, "AB̃CD" and "EB̃CD", were presented to the network alternately. Here, B̃ is a noisy version of pattern "B" (*Methods*). (*C*, *Left*) Mean network responses to sequence "ABCD" over different days. (*C*, *Right*) Same as the left figure, but for sequence "EB̃CD. (*D*) Average firing rates in the early and late phases, following the onset of the second stimulus "B̃" preceded by "A" (*Left*) or "E" (*Right*) over different days are shown. (*E*) Activity difference between day 4 and day 1 for the two sequences during early and late phases are shown. Consistent with an experiment by Price et al. (31), only the late phase response to "AB̃" was suppressed significantly through learning. (*F*) Same as *D*, but for average total recurrent synaptic currents in assembly B. (*G*) Same as *E*, but for average total recurrent synaptic currents in assembly B. In *D*–*G*, *P*-values were calculated using two-sided Welch's *t* tests (*$P < 0.05$, **$P < 0.01$). Data points for each case were generated by 10 independent simulations.

stimulating assembly "B" most strongly (*Methods*). To quantify to what extent these stimulation protocols influenced network activity, we split the period during which "B̃" was presented into an early (1 to 50 ms) and a late phase (51 to 100 ms) (Fig. 4*C*). It should be noted that our definitions of early and late phases are shifted up by 50 ms compared to those used in the mouse V1 experiment (i.e., 51 to 100 ms for early and 101 to 150 ms for late phase) (31).

Using this stimulus design, we calculated the average network activity observed in response to the expected and unexpected sequences in the early and late phases for each day. Similar to Price et al. (31)'s findings, neuronal responses to "B̃" changed from the pre- to postlearning stages, tending to decrease for both sequence types in the late phase. For both sequences, the decreases in late phase responses were significant, with the suppression measured for the "AB̃" sequence being significantly larger than that measured for the "EB̃" sequence (Fig. 4*E*). This is broadly congruent with the experimental data in which both sequences showed a decrease trend, but only the decrease in the "AB̃" sequence response was significant (Fig. 4*D*). Furthermore, as in the experimental data, no significant change was observed for the early phase, unless definitions of the early and late phases that did not take visual processing delays into account were used (*SI Appendix*, Fig. S9). Thus, our findings largely recapitulate the experimental results reported by Price et al. (31). It should be noted, however, that the specific experimental result replicated here is preliminary, as it was included in Price et al.'s preprint on bioRxiv (31), but not their final published version (38). Based on our model analyses, however, we predict that a follow-up study would confirm these preliminary findings.

To explore the origin of these experience-dependent prediction error signals, we calculated the mean recurrent synaptic currents within each assembly in response to the expected and unexpected sequences in the early and late phases for each day (Fig. 4*F*). Similar to the patterns observed when analyzing whole network responses, while the strength of synaptic currents decreased over training days for both sequence types, the suppression measured for the "AB̃" sequence was significantly larger than that measured for the "EB̃" sequence (Fig. 4*G*). We further found that, in the late phase, the differences in excitatory or inhibitory synaptic currents alone between day 4 and day 1 for the two sequences were not statistically significant (*SI Appendix*, Fig. S10 *B* and *D*; Late). Interestingly, while the change in total synaptic current did not show a significant difference in the early phase (Fig. 4*G*; Early), both excitatory and inhibitory currents separately showed significant changes (*SI Appendix*, Fig. S10 *B* and *D*; Early).

In summary, as we have shown, our model can explain how, through training, neuronal assemblies learn to respond differently to stimuli resulting from predictable transitions compared to unpredictable ones. Specifically, our model successfully recapitulates experimental results from the mouse primary visual cortex in which late phase, but not early phase, neural responses are suppressed with experience following an expected stimulus transition, but not an unexpected one.

Learning Context-Dependent Prediction Error Signals. So far, we have demonstrated that our network can develop a comparatively simple class of prediction error-related activity. Indeed, in the simulations so far, prediction errors were primarily generated by a violation of a particular transition between a pair of stimuli. Although it is possible that prediction errors only encode a generic error signal, shared across multiple stimuli, a recent experimental study showed that prediction error signals for particular stimuli emerge in a context-dependent manner (30). This experiment demonstrated that many neurons that showed strong suppression in response to specific stimuli only did so in the expected context. This suggests that a highly selective network mechanism exists for encoding context-dependent expectations. It follows that nonspecific and population-wide inhibition of excitatory neurons that encode specific stimuli cannot on its own explain the emergence of expectation-based signals. Instead, inhibitory connections may be precisely tuned to only generate expectations for sensory stimuli in selective contexts. However, a plasticity mechanism that could explain how context-dependent prediction errors emerge is still elusive. We therefore wondered whether our model could account for learning such context-dependent prediction error representations.

In Audette & Schneider (30), mice were trained to perform a sound-generating spontaneous forelimb movement task to explore how movement-based predictions affect neural responses to expected and unexpected sounds. During training, a stereotypical auditory stimulus was presented each time mice generated a forelimb movement. After the animal had undergone sufficient training to spontaneously produce these movements, mice heard either the well-trained expected sound or a novel auditory stimulus with a slightly different frequency at the beginning of each forelimb movement they produced ("active" condition). These sounds were also played in conditions where the animal was not performing forelimb movements ("passive" condition). The experiment showed that neural responses to the expected sound in the active condition were suppressed compared to the same sound heard in the passive case. In contrast, responses to the unexpected sound under the active condition were enhanced relative to the passive case. This result suggests that prediction error signals emerge due to a specific combination of stimulus and context in a way that is dependent on expectation.

We show that our model can learn context-specific prediction errors as well. To this end, we considered two types of inputs: one corresponding to auditory signals, and the other to a motor signal from the motor cortex. Excitatory populations were divided into two distinct populations, each of them receiving one auditory signal (i.e., familiar or novel; A or B) (Fig. 5 *A*, *Top*). All neurons, excitatory and inhibitory, received distributed step-shaped inputs to model motor command signals. We assumed that presentation of both auditory and motor command inputs increased excitatory drive to neurons targeted by each pattern. During learning, the auditory signals A and B were presented to the network alternately, with only signal A being combined to the motor signal (active condition, Fig. 5 *A*, *Bottom*).

After learning, we first compared the network responses to the familiar stimulus (i.e., stimulus pattern "A") under the active and the passive cases. We simulated responses under the active condition by measuring evoked dynamics in the network receiving both stimulus pattern A and a motor command signal (Fig. 5 *B*, *Left*). In contrast, we defined responses to stimulus A under the passive condition as network responses to stimulus pattern A alone (Fig. 5 *B*, *Right*). Consistent with the experimental results, the network responses to the expected sound in the active condition were suppressed compared to the responses to the same sound heard in the passive case (Fig. 5 *C*, *Left*). To test whether our model showed context-dependent prediction errors as shown experimentally (30), we also compared the network responses to the stimulus pattern B under the active and passive cases. Interestingly, in contrast to stimulus A, responses to the stimulus B in the active condition were enhanced relative to the passive case (Fig. 5 *C*, *Right*).

We then asked what the potential mechanism is underlying the emergence of context-dependent prediction errors in the network model. As we saw in the simple task, prediction error signals are generated by breaking the EI balance with an unexpected stimulus presentation. We therefore monitored excitatory and inhibitory

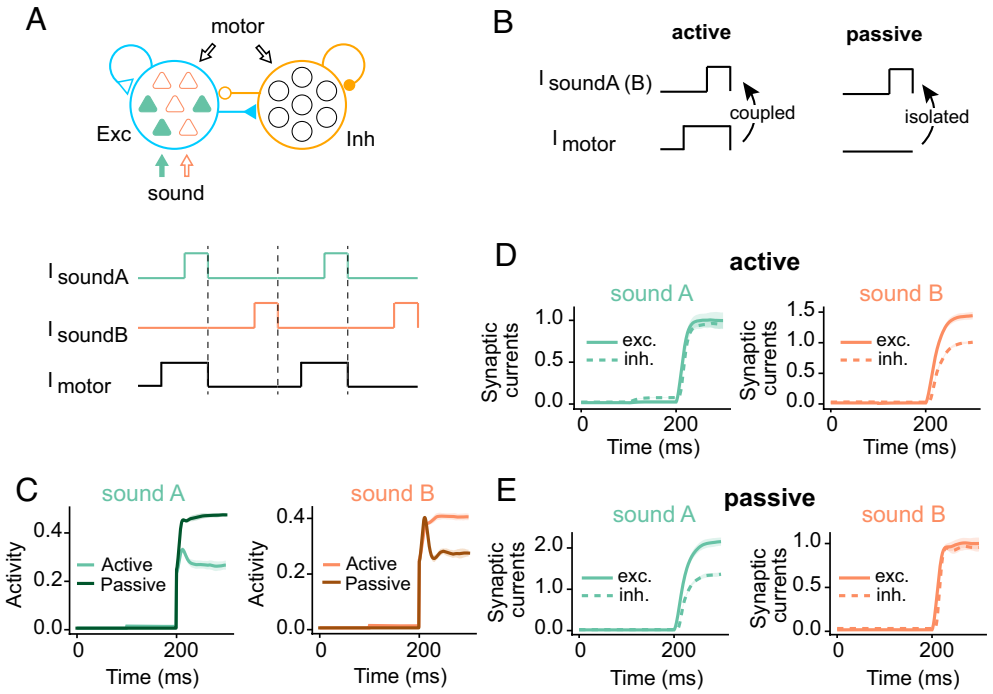

Fig. 5. Learning of context-dependent prediction error signals. (*A*, *Top*) Model schematic. Excitatory population was divided into two subpopulations, each of which received either sound stimulus A or B (green and orange arrows). In addition, all neurons in the network received motor command input. (*A*, *Bottom*) During learning, the auditory signal A was combined to the motor signal, but signal B was isolated from any motor signal. (*B*) After learning, sounds A and B were presented either coupled with (active) or isolated from (passive) the motor input. (*C*) Mean network responses to sounds A and B in the passive (Darker) and active (Lighter) context. (*D*) Recurrent excitatory and inhibitory currents while sound A (*Left*) or B (*Right*) was presented in the active context are shown. (*E*) Same as *D*, but in the passive context.

recurrent currents during stimulus pattern A or B, as presented under two different conditions (i.e., active or passive condition). When stimulus A was presented, the strength of both currents was not drastically different in the active case (Fig. 5 *D*, *Left*), but was significantly different in the passive condition (Fig. 5 *E*, *Left*). In contrast, when the novel pattern B was presented, the EI balance was maintained in the passive case (Fig. 5 *E*, *Right*), but broken in the active condition (Fig. 5 *D*, *Right*).

In summary, these results suggest that our network model with prediction-based plasticity can learn context-dependent prediction errors, as shown in the experimental study of Audette & Schneider (30). The results also showed that prediction errors were generated due to a disruption of the EI balance that was precisely tuned in a context-dependent manner. Although we have framed this here as contextual modulation based upon a signal from the motor cortex, the same set-up could apply to other (passive) forms of contextual modulation (e.g., a visual stimulus that conditions expectations as to the likely auditory stimulus). This is important to note as there may be additional specific mechanisms in play in the motor system including the sensory attenuation that can emerge following learning in active settings (39).

Discussion

In this study, we investigated how plasticity at recurrent excitatory and inhibitory synapses can produce prediction errors that carry features of mismatch responses observed in the brain. Specifically, we trained our model to predict upcoming network activity through its excitatory synapses, while its inhibitory synapses were tuned to maintain the EI balance (29). We showed that the network learned the appropriate connectivity patterns to encode stimulus statistics and generate prediction error signals when unexpected stimuli were presented, in agreement with various experimental results (2, 5–9, 30, 31).

Predictive coding suggests that mismatch signals may carry prediction errors in the brain. Indeed, previous computational studies have shown that spiking network models with layered cortical architecture trained using predictive coding replicate mismatch signals (3). Notably, however, although both predictive coding models and our recurrent network model generate prediction errors, there are several differences between the two. First, in traditional predictive coding models, predicted state and prediction error signals are typically encoded in separate populations of neurons (18–20, 40–42). In contrast, in our model, both the predicted states and prediction errors are represented within a single neuron. We achieved this by implementing two distinct nonlinear activation functions within single neurons. A potential biologically plausible implementation of this feature would be to explicitly implement neurons as two-compartment units, where a dendritic compartment is nonlinearly connected to a somatic compartment, which produces the neuron's output. We leave the question of how exactly prediction and error signals might be encoded in more biologically plausible segregated neuron compartments and the associated plasticity rules to future work.

Another difference between traditional predictive coding and our model is that, in predictive coding, top–down input predicts bottom–up input (18–20, 43). In contrast, in our model, predictions are generated locally, by the recurrent input. Experiments have shown that mismatch responses occur even when participants are engaged in a distractor task that draws attention away from the sensory modality in which the oddball stimulus occurs. This suggests that top–down input may not always be necessary for generating prediction error signals (2, 44). An alternative interpretation is that the locally computed predictions act as a distributed top–down input to be compared to the bottom–up signal from stimulus presentation. It is this comparison that leads to the (again, distributed) prediction error signal measurable through population averages.

We showed that local recurrent connectivity is sufficient to reproduce different kinds of prediction error signals. Although a previous study suggested that combining recurrent connectivity with top–down prediction supports associative memory tasks via covariance learning, in that study, recurrent connections were trained to predict specifically in the spatial domain (27). Due to the nature of recurrent plasticity, the direct relationship between recurrent input and prediction error signals generated over time has remained elusive. We found that our model learns predictions in the temporal domain via local recurrent circuit and thus generates prediction error signals over sequential stimuli. How prediction error signals can encode both spatial and temporal information remains an open question.

Our model reproduces the general temporal profile of mismatch responses. Experimentally, several types of prediction error signals, including the standard MMN and inferotemporal cortical mismatch responses to visual stimuli violating sequence transitional rules, show a biphasic waveform (8, 9). This waveform typically consists of an early negative deflection, followed by a positive one. Although one must be cautious in interpreting the meaning of positive and negative waves in EEG studies, it is notable that a very similar biphasic pattern emerges when measuring the difference in overall neural activity in response to expected and unexpected stimuli in our recurrent network.

Predictive coding in which the prediction and prediction error signals are carried by different neurons (18–20, 40, 41), as opposed to a single neuron as in our study, may present some benefits. In particular, this may allow for more complex error signals to emerge through populations of error neurons specializing, for example, in positive and negative errors, respectively. Hertäg and Clopath (24) show that such a network can, for example, use negative prediction error neurons to represent when an actual sensory stimulus is smaller than predicted, and positive prediction error neurons to represent when an actual sensory stimulus is bigger than predicted (24). In our model, it is conceivable that a similar function could be achieved using a temporal code, i.e., by modulating the amplitude of the positive or negative components of the prediction error signal, specifically. Indeed, we showed that both negative and positive prediction errors could arise from a breaking of the EI balance. Specifically, negative prediction errors were due to a significant decrease in excitatory currents, whereas positive ones were due to a decrease in inhibitory currents. Further study is required to determine what advantages and disadvantages these different implementations present, and which best explains the spatial and temporal properties of mismatch signals in the brain.

The biphasic prediction error dynamics could also be interpreted in the context of probabilistic inference, where a Markov chain augmented with temporal smoothing could exhibit similar behavior. Specifically, a forward inference mismatch (e.g., at the unexpected transition) could generate a negative signal, followed by a smoothing process on a slower timescale, resulting in a subsequent positive phase. While this interpretation provides a possible algorithmic explanation, our work provides a complementary perspective by showing how such dynamics can emerge mechanistically from biologically plausible synaptic plasticity in recurrent networks. Furthermore, beyond the learning of prediction error signals, recent studies on predictive inference—particularly in the context of sequence learning with context-dependent latent learning—have demonstrated that latent variable models can effectively capture complex prediction signals (45, 46). These approaches employ Expectation–Maximization algorithms, which have been connected to spike-timing-dependent plasticity for sequence learning (47). Further experimental studies are needed to bridge the gap between abstract probabilistic inference algorithms and the dynamics of prediction error signals in neural circuits.

Although our model can account for learning a variety of prediction error signals, in principle, it cannot learn predictions over the global structure of stimulus chunks. The reason behind this shortcoming is that our model learns local transition statistics between stimuli only, but is not designed to learn higher-order statistics (e.g., non-Markovian statistics) of stimuli. There are several possible ways to overcome this limitation. One possible solution is to consider much longer time scales than we considered in this study, such as calcium dynamics or N-methyl-d-aspartate receptor-mediated (NMDA) spikes. Indeed, a previous computational study showed that a NMDA-dependent plasticity-like rule could enable prediction error signals to be learned over the global structure of a sequence (3). Another possible way to achieve this is to consider hierarchically structured networks, similar to real cortical regions. In such hierarchical networks, subnetworks lower in the hierarchy could learn to encode local element-level transitions, while the higher-level networks could learn slow and abstract dynamics (48), thus developing error signals related to the global structure of the sensory inputs. Extending our recurrent network model into a hierarchically structured model, and studying the relationship between recurrently driven and top–down-driven prediction error signals could shed important light on the difference between global and local mismatch signals in the brain.

In conclusion, our study sheds light on the learning mechanisms that may underlie mismatch signals in the brain, and provides a perspective on the relationship between synaptic plasticity and prediction errors. Furthermore, it opens up broad avenues for future studies of prediction error signals in hierarchical networks, and may contribute to the development of more flexible and biologically plausible models of neural computation.

Methods

Our recurrent neural networks consist of N_E excitatory and N_I inhibitory neurons. During learning, the membrane potentials of neuron i at time t with external current I_i^{ext} were calculated as follows:

$$u_i^E(t) = \sum_{j=1}^{N_E} W_{ij}^{EE} x_j^E(t) - \sum_{k=1}^{N_I} W_{ik}^{EI} x_k^I(t) + I_i^{ext}(t), \quad [1]$$

$$u_i^I(t) = \sum_{j=1}^{N_E} W_{ij}^{IE} x_j^E(t) - \sum_{k=1}^{N_I} W_{ik}^{II} x_k^I(t), \quad [2]$$

where u_i^E and u_i^I are the membrane potential of i-th excitatory and inhibitory neuron, respectively (see Table 1 for the list of variables and functions). The strength of external input I_i^{ext} takes the value 1 if a stimulus pattern targeting neuron i was presented, and 0 otherwise. This structured external input was replaced by constant inputs I_i^{const} of value 0.3 during spontaneous activity. We will describe the details of stimulus patterns later. $W_{ij}^{ab}\left(a, b = E; I\right)$ is a recurrent

Table 1. Definition of variables and functions

u_i^E, u_i^I	Membrane potentials
x_j^E, x_k^I	Postsynaptic potentials
S_i^a	Poisson spike train generated by network neurons
W_{ij}^{EE}, W_{ik}^{EI}, W_{ij}^{IE}, W_{ik}^{II}	Recurrent connections
I_i^E, I_i^I	Synaptic currents generated by network neurons
f_i^E, f_i^I	Instantaneous firing rates
y_i^E, y_i^I	Recurrent predictions
φ	Sigmoidal function

connection weight from j-th neuron in population b to i-th neuron in population α. All neurons were connected with a coupling probability of $P = 0.5$. The initial values of synaptic weights W_{ij}^{ab} were uniformly set to $0.5/\sqrt{pN_b}$). x_i^a is a postsynaptic potential evoked by i-th neuron in population a, which will be described later.

Spiking of each neuron model in population E was modeled as an inhomogeneous Poisson process with instantaneous firing rate f_i^E with a sigmoidal response function φ, with parameters slope β and threshold θ, as:

$$f_i^E = \varphi\left(u_i^E\right) \equiv \varphi_0\left[1+\exp\left[\beta\left(-u_i^E+\theta\right)\right]\right]^{-1}, \quad [3]$$

where φ_0 is the maximum instantaneous firing rate of 50 Hz. Throughout the figures in this paper, the normalized firing rate (i.e., f_i^E/φ_0) was simply referred to as the firing rate.

Inhibitory neurons' firing rates were assumed to be calculated with static sigmoidal function as:

$$f_i^I = \varphi\left(u_i^I\right) \equiv \varphi_0\left[1+\exp\left[\beta\left(-u_i^I+\theta\right)\right]\right]^{-1}, \quad [4]$$

where the maximum instantaneous firing rate φ_0 was assumed to be the same as that of excitatory neurons (i.e., 50 Hz).

Neuron i in population a generates a Poisson spike train at the instantaneous firing rate of f_i^a. Let us describe the generated Poisson spike trains as:

$$S_i^a(t) = \sum_{t' \in t_i^a} \delta\left(t-t'\right), \quad [5]$$

where δ is Dirac's delta function and t_i^a is the set of times at which a spike occurred in the neuron. The postsynaptic potential evoked by the neuron (i.e., x_i^a) is then calculated as:

$$\tau_s \dot{I}_i^a = -I_i^a + \frac{1}{\tau} S_i^a, \quad [6]$$

$$\dot{x}_i^a = \frac{-x_i^a}{\tau} + x_0 I_i^a, \quad [7]$$

where $\tau_s = 5$ ms, $\tau = 15$ ms, and $x_0 = 25$.

The Learning Rules. As in ref. 29, all excitatory synaptic connections onto excitatory neurons obeyed the following plasticity rule to predict the activity of postsynaptic neurons as:

$$\Delta W_{ij}^{EE} = \epsilon\left[f_i^E - y_i^E\right] \bullet x_j^E, \quad [8]$$

where y_i^E is a recurrent prediction of a firing rate, defined as:

$$y_i^E = \varphi\left(\sum_{j=1}^{N_E} W_{ij}^{EE} \bullet x_j^E\right), \quad [9]$$

where the function $\varphi(\bullet)$ is the sigmoid function defined in Eq. **3**. In this study, the learning rate was set to $\epsilon = 10^{-4}$ in all simulations.

The inhibitory synapses onto excitatory neurons were plastic according to the following rule:

$$\Delta W_{ij}^{EI} = \epsilon\left[y_i^E - y_i^I\right] x_j^I, \quad [10]$$

where y_i^I was the total inhibitory input onto postsynaptic neuron:

$$y_i^I = \varphi\left(\sum_{j=1}^{N_I} W_{ij}^{EI} \bullet x_j^I\right). \quad [11]$$

Through this inhibitory plasticity, inhibitory synapses were modified to maintain excitatory-inhibitory balance in all excitatory neurons.

Simulation Details. The parameters used in the simulations are summarized in Table 2. All simulations were performed in customized Python3 code written by TA with numpy 1.17.3 and scipy 0.18. Differential equations were numerically integrated using a Euler method with integration time steps of 1 ms.

Table 2. Parameter settings

p	Connection probability	0.5
N_E, N_I	Network size	500
ϵ	Learning rate	10^{-4}
τ_s	Synaptic time constant	5 ms
τ	Membrane time constant	15 ms
β, θ	Parameters for sigmoid	5, 1
φ_0	Maximal firing rate	50 Hz
x_0	Scaling factor of synaptic current	25
I_i^{ext}	External current elicited by stimulus presentation	1
I_i^{const}	Constant external current during testing phase	0.3

Stimulation Protocols. In all simulations, each stimulus pattern had a duration of 100 ms and was presented in sequence without an interpattern interval, except in Fig. 4, where the duration was set to 150 ms. Sequences of patterns were repeatedly presented at 300-ms intervals; when multiple sequences were used, they were shown alternately at the same interval. We assumed each neuron in a network was stimulated by one of the stimulus patterns. Presentation of each pattern triggered excitatory currents to its targeted neurons of strength 1 and zero otherwise. In all simulations, we assumed that only external inputs caused by the presentation of the stimuli were injected into the network during learning. In contrast, we assumed all excitatory neurons received both structured and constant background inputs over the whole period occurring after learning. During learning, all excitatory synaptic connections onto excitatory neurons were assumed to be plastic, while they were static during the testing phase after learning. The network was trained typically for 1,000 s except in Fig. 4, where the simulation time was 300 s.

Measuring Prediction Error Signals. In Fig. 2*D*, the early and late phases of responses were defined as the periods before and after the point at which the mean expected and unexpected responses intersected. Mean responses were calculated over 10 independent simulations. We then calculated the mean differences between responses over unexpected and expected sequences (unexpected-expected) for the two periods (i.e., early and late phases).

Learning with Overlapping Assemblies. In *SI Appendix*, Fig. S3, the excitatory population was divided into three cell assemblies. Each assembly corresponded to a specific stimulus, and shared subsets of neurons represented the overlap between stimulus pairs. Specifically, neurons shared between any two assemblies corresponded to neurons associated with the respective stimulus pair. The proportion of neurons shared between each pair of assemblies was set to 60% of the total population. The remaining neurons were uniquely assigned to individual assemblies and were not shared between groups.

Evaluation of Prediction Performance. In *SI Appendix*, Figs. S3*C* and S4*C*, state transition prediction performance was evaluated using a spiking recurrent network. In both cases, averaged recurrent inputs were converted to prediction probabilities using a softmax function (coefficient of 10), and performance was measured using negative log-likelihood. In *SI Appendix*, Fig. S3*C*, every 3 s during learning, while the cell assembly corresponding to state 2 was driven, we computed the population average of recurrent inputs to the nonoverlapping parts of the cell assembly corresponding to state 3. To eliminate self-transitions, we set the input to assembly 2 to zero. We ran 20 independent simulations to measure the average model performance. In *SI Appendix*, Fig. S4*C*, the same procedure was applied to transitions from state 4; however, since state 4 can transition to both state 2 and state 3, prediction probabilities were computed for both transitions, and the negative log likelihood was determined using the average of these probabilities. For comparison with the performance of a machine learning method, we also measured prediction performance using a clone-structured cognitive graph method (45). The pseudocount was set to 10^{-30} and we used 1 clone.

Learning of Expectation-Dependent Prediction Error Signals. In Fig. 4 *C* and *D*, periods during which stimulus pattern "B" was presented were divided into an early and a late phase. Mean activities over time were calculated over different days for each condition. To simulate stimulation with noisy stimulus "$\tilde{B}$," assembly B was stimulated with an input of intensity 0.6 I^{ext} and the two other assemblies (i.e., assemblies A and D) with an input of intensity 0.1 I^{ext}, where I^{ext} is the strength of the input for all other patterns (i.e., A, B, "C", and "D"). Specifically, the additional stimulation of assemblies A and D during "$\tilde{B}$" is intended to capture the experimental effect of introducing jitter into the orientation of stimulus B.

Data, Materials, and Software Availability. The source code for reproducing the main figures is available on GitHub: https://github.com/TAsabuki/prediction_error (49). All other data are included in the manuscript and/or supporting information.

ACKNOWLEDGMENTS. This work was supported by Biotechnology and Biological Sciences Research Council (BB/N013956/1), Wellcome Trust (200790/Z/16/Z), the Simons Foundation (564408), and Engineering and Physical Sciences Research Council (EP/R035806/1). We are grateful to Loreen Hertäg for fruitful discussions.

1. G. B. Keller, T. D. Mrsic-Flogel, Predictive processing: A canonical cortical computation. *Neuron* **100**, 424–435 (2018).
2. M. I. Garrido, J. M. Kilner, K. E. Stephan, K. J. Friston, The mismatch negativity: A review of underlying mechanisms. *Clin. Neurophysiol.* **120**, 453–463 (2009).
3. C. Wacongne, J. P. Changeux, S. Dehaene, A neuronal model of predictive coding accounting for the mismatch negativity. *J. Neurosci.* **32**, 3665–3678 (2012).
4. G. B. Keller, T. Bonhoeffer, M. Hübener, Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* **74**, 809–815 (2012).
5. E. Liebenthal *et al.*, Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *Neuroimage* **19**, 1395–1404 (2003).
6. S. J. Luck, *An Introduction to the Event-related Potential Technique* (MIT press, 2014).
7. C. Wacongne *et al.*, Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20754–20759 (2011).
8. T. Nagai *et al.*, Reduced mismatch negativity is associated with increased plasma level of glutamate in first-episode psychosis. *Sci. Rep.* **7**, 2258 (2017).
9. T. Meyer, C. R. Olson, Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc. Nat. Acad. Sci. U.S.A.* **108**, 19401–19406 (2011).
10. J. M. Strömmer, I. M. Tarkka, P. Astikainen, Somatosensory mismatch response in young and elderly adults. *Front. Aging Neurosci.* **6**, 293 (2014).
11. M. Kimura, J. Katayama, H. Ohira, E. Schröger, Visual mismatch negativity: New evidence from the equiprobable paradigm. *Psychophysiology* **46**, 402–409 (2009).
12. M. Kimura, E. Schröger, I. Czigler, H. Ohira, Human visual system automatically encodes sequential regularities of discrete events. *J. Cognit. Neurosci.* **22**, 1124–1139 (2010).
13. I. Winkler, I. Czigler, E. Sussman, J. Horváth, L. Balázs, Preattentive binding of auditory and visual stimulus features. *J. Cognit. Neurosci.* **17**, 320–339 (2005).
14. R. Näätänen, K. Alho, Higher-order processes in auditory-change detection. *Trends Cognit. Sci.* **1**, 44–45 (1997).
15. E. Schröger, C. Wolff, Mismatch response of the human brain to changes in sound location. *Neuroreport* **7**, 3005–3008 (1996).
16. B. Opitz, T. Rinne, A. Mecklinger, D. Y. von Cramon, E. Schröger, Differential contribution of frontal and temporal cortices to auditory change detection: FMRI and ERP results. *Neuroimage* **15**, 167–174 (2002).
17. R. Näätänen, T. Jacobsen, I. Winkler, Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology* **42**, 25–32 (2005).
18. A. M. Bastos *et al.*, Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711 (2012).
19. R. P. Rao, D. H. Ballard, Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
20. Y. Huang, R. P. Rao, Predictive coding. *Wiley Interdis. Rev. Cognit. Sci.* **2**, 580–593 (2011).
21. F. Lieder, J. Daunizeau, M. I. Garrido, K. J. Friston, K. E. Stephan, Modelling trial-by-trial changes in the mismatch negativity *PLoS Comput. Biol.* **9**, e1002911 (2013).
22. A. Fiser *et al.*, Experience-dependent spatial expectations in mouse visual cortex. *Nat. Neurosci.* **19**, 1658–1664 (2016).
23. M. Garrett *et al.*, Experience shapes activity dynamics and stimulus coding of VIP inhibitory cells. *Elife* **9**, e50340 (2020).
24. L. Hertäg, C. Clopath, Prediction-error neurons in circuits with multiple neuron types: Formation, refinement, and functional implications. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2115699119 (2022).
25. K. Friston, I. Herreros, Active inference and learning in the cerebellum. *Neural Comput.* **28**, 1812–1839 (2016).
26. B. Millidge *et al.*, A theoretical framework for inference and learning in predictive coding networks. arXiv [Preprint] (2023). https://arxiv.org/abs/2207.12316 (Accessed 13 June 2023).
27. M. Tang *et al.*, Recurrent predictive coding models for associative memory employing covariance learning. *PLoS Comput. Biol.* **19**, e1010719 (2023).
28. Y. Song *et al.*, Inferring neural activity before plasticity as a foundation for learning beyond backpropagation. *Nat. Neurosci.* **27**, 348–358 (2024).
29. T. Asabuki, C. Clopath, Embedding stochastic dynamics of the environment in spontaneous activity by prediction-based plasticity. *Elife* **13**, RP95243 (2025).
30. N. J. Audette, D. M. Schneider, Stimulus-specific prediction error neurons in mouse auditory cortex. *J. Neurosci.* **43**, 7119–7129 (2023).
31. B. H. Price, C. M. Jensen, A. A. Khoudary, J. P. Gavornik, Expectation violations produce error signals in mouse V1. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2021.12.31.474652 (Accessed 13 June 2023).
32. J. P. Pfister, T. Toyoizumi, D. Barber, W. Gerstner, Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural Comput.* **18**, 1318–1348 (2006).
33. R. Urbanczik, W. Senn, Learning by the dendritic prediction of somatic spiking. *Neuron* **81**, 521–528 (2014).
34. T. Asabuki, T. Fukai, Somatodendritic consistency check for temporal feature segmentation. *Nat. Commun.* **11**, 1554 (2020).
35. T. Asabuki, T. Fukai, Predictive learning rules generate a cortical-like replay of probabilistic sensory experiences. *Elife* **13**, RP92712 (2024).
36. R. Näätänen, A. W. Gaillard, S. Mäntysalo, Early selective-attention effect on evoked potential reinterpreted. *Acta Psycholog.* **42**, 313–329 (1978).
37. M. Strauss *et al.*, Disruption of hierarchical predictive coding during sleep. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1353–E1362 (2015).
38. B. H. Price, C. M. Jensen, A. A. Khoudary, J. P. Gavornik, Expectation violations produce error signals in mouse V1. *Cereb. Cortex.* **33**, 8803–8820 (2023).
39. H. Idei, W. Ohata, Y. Yamashita, T. Ogata, J. Tani, Emergence of sensory attenuation based upon the free-energy principle. *Sci. Rep.* **12**, 14542 (2022).
40. K. Friston, Prediction, perception and agency. *Int. J. Psychophysiol.* **83**, 248–252 (2012).
41. K. Friston, Does predictive coding have a future? *Nat. Neurosci.* **21**, 1019–1021 (2018).
42. S. Shipp, Neural elements for predictive coding. *Front. Psychol.* **7**, 1792 (2016).
43. W. Lotter, G. Kreiman, D. Cox, Deep predictive coding networks for video prediction and unsupervised learning. arXiv [Preprint] (2016). https://arxiv.org/abs/1605.08104 (Accessed 13 June 2023).
44. R. Näätänen, P. Paavilainen, T. Rinne, K. Alho, The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clin. Neurophysiol.* **118**, 2544–2590 (2007).
45. D. George *et al.*, Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. *Nat. Commun.* **12**, 2392 (2021).
46. R. V. Raju *et al.*, Space is a latent sequence: A theory of the hippocampus. Science. *Sci. Adv.* **10**, eadm8470 (2024).
47. D. Kappel, B. Nessler, W. Maass, STDP installs in winner-take-all circuits an online approximation to hidden Markov model learning. *PLoS Comput. Biol.* **10**, e1003511 (2014).
48. A. Maes, M. Barahona, C. Clopath, Learning compositional sequences with multiple time scales through a hierarchical network of spiking neurons. *PLoS Comput. Biol.* **17**, e1008866 (2021).
49. T. Asabuki, Code for simulations from "Learning predictive signals within a local recurrent circuit". GitHub. https://github.com/TAsabuki/prediction_error. Deposited 13 June 2025.