Contextual inference underlies the learning of sensorimotor repertoires

James B. Heald[1,2,*], Máté Lengyel[2,3‡] & Daniel M. Wolpert[1,2‡]

[1] Zuckerman Mind Brain Behavior Institute, Department of Neuroscience, Columbia University, NY, USA

[2] Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Cambridge, UK

[3] Center for Cognitive Computation, Department of Cognitive Science, Central European University, Budapest, Hungary

‡These authors contributed equally to this work.

*corresponding author

Humans spend a lifetime learning, storing and refining a repertoire of motor memories. For example, through experience, we become proficient at manipulating a large range of objects with distinct dynamical properties. However, it is unknown what principle underlies how our continuous stream of sensorimotor experience is segmented into separate memories and how we adapt and use this growing repertoire. Here we develop a theory of motor learning based on the key principle that memory creation, updating and expression are all controlled by a single computation – contextual inference. Our theory reveals that adaptation can arise both by creating and updating memories (proper learning) and by changing how existing memories are differentially expressed (apparent learning). This insight allows us to account for key features of motor learning that had no unified explanation: spontaneous recovery[1], savings[2], anterograde interference[3], how environmental consistency affects learning rate[4,5] and the distinction between explicit and implicit learning[6]. Critically, our theory also predicts novel phenomena – evoked recovery and context-dependent single-trial learning – which we confirm experimentally. These results suggest that contextual inference, rather than classical single-context mechanisms[1,4,7–9], is the key principle underlying how a diverse set of experiences is reflected in our motor behaviour.

Throughout our lives, we experience different contexts, in which the environment exhibits distinct dynamical properties, such as when manipulating different objects or walking on different surfaces. Although it has been recognised that the brain maintains multiple motor memories appropriate for these contexts[10,11], classical theories of motor learning have focused on how the brain adapts to a single type of environmental dynamics[1,7,8]. However, with multiple memories come new computational challenges: the brain must decide when to create new memories[12] and how much to express and update them for each movement we make. These operations, their governing principles and consequences on motor learning, remain poorly understood. Here, we propose a unifying principle – contextual inference – that specifies how sensory cues and state feedback affect memory creation, expression and updating. We show that contextual inference is the core feature that underlies a range of fundamental aspects of motor learning that were previously explained by a number of distinct and often heuristic processes.

COIN: a model of contextual inference

In order to formalise the role of contextual inference in motor learning, we developed the COIN (COntextual INference) model, a principled nonparametric Bayesian model of motor learning (see Methods). The COIN model is based on an internal model that specifies the learner's assumptions about how the environment generates their sensory observations (Fig. 1a, Extended Data Fig. 1a). Motor learning corresponds to online Bayesian inference under this generative model (Fig. 1b, Extended Data Fig. 1b). For this, the COIN model jointly infers contexts, their transitions, their dynamical and sensory properties, and the current state of each context, such that each motor memory stores the learner's inferences about a different context (for validation, see Extended Data Fig. 2a-b). The major challenge in motor learning is that neither contexts nor their transitions come labelled, and thus the learner needs to continually infer which context they are in based on a continuous stream of experience.

The result of contextual inference is a posterior distribution expressing the probability with which each known context, or a yet-unknown novel context, is currently active (Fig. 1b, top row). In turn, contextual inference determines memory creation, expression and updating (Fig. 1b, numbered arrows). Fig. 1c-f (and Extended Data Fig. 1c-e) illustrates this in a simulation of the COIN model (parameters in Extended Data Fig. 3) when handling objects of varying weights. For determining the current motor command (Fig. 1e), rather than selecting a single memory to be expressed[11,12], the state associated with each memory (Fig. 1d) is expressed commensurate with the probability of the corresponding context under the posterior, computed after observing the sensory cue but before movement ('predicted probability'; Fig. 1b, arrow 1; Fig. 1f_1). After movement, the 'responsibility' of each known context as well as of a novel, yet-unknown context is computed as their posterior probability given both the cue and the resultant state feedback. A new memory is created flexibly, whenever the responsibility of a novel context becomes high (Fig. 1b, arrow 2; Fig. 1f_2). Critically, context responsibilities also scale the updating of the previously existing memories and any newly created memory (Fig. 1b, arrows 3; Fig. 1f_3, red and pink arrows respectively showing how high and low responsibility for the red context speeds up and slows down the updating of its state, Fig. 1d). Finally, these responsibilities are used to compute the predicted context probabilities on the next time step (Fig. 1f_1).

In summary, the COIN model proposes that contextual inference is core to motor learning. In particular, unlike in traditional models of learning, adaptation to a change in the environment (e.g. Fig. 1e, blue and cyan arrows) can arise from two distinct and interacting mechanisms. First, in line with classical notions of learning, *proper* learning constitutes the creation and updating of memories (the inferred states of known contexts; Fig. 1d, blue arrow). Second, *apparent* learning occurs due to the updating of the predicted context probabilities (Fig. 1f_1, cyan arrow), thereby altering the extent to which existing memories are ultimately expressed in behaviour.

Apparent learning underlies memory recovery

As an ideal litmus test of the contributions of contextual inference to memory creation and expression (Fig. 1b, arrows 1-2), we revisited a widely-used motor learning paradigm. In this paradigm (Fig. 2a and b, top left), participants learn a perturbation P^{+} applied by a robotic interface while reaching to a target. Adaptation is assessed using occasional channel trials, P^{c}, which remove movement errors and measure the forces participants use to counteract the perturbation (Fig. 2a, see Methods for details). Exposure to P^{+} is followed by brief exposure to the opposite perturbation, P^{-}, bringing adaptation back, near to baseline. Finally, a series of channel trials is administered. As in previous studies[1], our participants showed the intriguing feature of spontaneous recovery in this phase (Fig. 2c): a transient re-expression of P^{+} adaptation, rather than a simple decay towards baseline.

Although this paradigm has no explicit sensory cues, according to our theory, contextual inference plays an important role. When simulated for this paradigm (Fig. 2b), the COIN model starts with a memory appropriate for moving in the absence of a perturbation (P^{0}, blue Fig. 2b, bottom left) and creates new memories for the P^{+} (red) and P^{-} (orange) perturbations. Spontaneous recovery arises due to the dynamics of contextual inference. As P^{+} has been experienced in most trials, it is quickly inferred to be active with a high probability during the channel-trial phase (Fig. 2b, top right). Therefore, as its state has not yet decayed (Fig. 2b bottom left), the memory of P^{+} is transiently expressed in the participant's motor output (Fig. 2b bottom right). This mechanism is fundamentally different from that of a classical, single-context model of motor learning, the dual-rate model[1]. There, motor output is determined by a combination of individual memories that update at different rates (fast and slow) but whose expression does not change over time. Thus the dynamics of adaptation is solely determined by the dynamics of memory updating, i.e. *proper* learning. In contrast, in the COIN model, changes in motor output can occur without updating any individual memory, simply due to changes in the extent to which existing memories are expressed due to contextual inference, i.e. *apparent* learning. This mechanism allows the COIN model to account robustly for spontaneous recovery (Extended Data Fig. 4a), including elevated or reduced levels when the P^{+} phase is extended[13] (Extended Data Fig. 5a-j) or when P^{-} is experienced prior to the P^{+} phase[14] (Extended Data Fig. 5k-o), respectively.

In order to distinguish between proper and apparent learning as the main mechanism underlying spontaneous recovery, we designed a novel 'evoked recovery' paradigm (similar to the reinstatement paradigm in classical conditioning[15]) in which sensorimotor evidence clearly indicates that a change in context has occurred. For this, two early trials in the channel-trial phase of the spontaneous recovery paradigm were replaced with P^{+} ('evoker') trials (Fig. 2d, top left, akin to trigger trials in visuomotor learning[11]). In this case, the COIN model predicts a strong and long-lasting recovery of P^{+}-adapted behaviour (Fig. 2d, bottom right; Extended Data Fig. 4b), primarily due to the inference that the P^{+} context is now active (Fig. 2d, top right, red) and the gradual decay of the P^{+} state over subsequent channel trials (Fig. 2d, bottom left, red). In addition, our mathematical analysis suggested that evoked as well as spontaneous recovery are inherent features of the COIN model (Suppl. Inf. and Extended Data Fig. 6a-c). In contrast, the dual-rate model only predicts a transient recovery that rapidly decays due to the same underlying adaptation process with fast dynamics governing both recovery and decay (Extended Data Fig. 6d).

In line with COIN model predictions, participants showed a strong evoked recovery in response to the P^{+} trials (Fig. 2e). This recovery lasted for the duration of the experiment, defying models that predict a simple exponential decay to baseline[4,11,16] (Extended Data Fig. 6e and Extended Data Table 1). We fit the COIN and dual-rate models to individual participants' data in both experiments (Fig. 2c & e). The COIN model fit the data accurately, but the dual-rate model (and its multi-rate extensions, Extended Data Fig. 6d) showed a qualitative mismatch in the time course of decay of evoked recovery (insets in Fig. 2c & e). Formal model comparison provided strong support for the COIN model overall (Δ group-level BIC of 302.6 and 394.1 nats for the spontaneous and evoked recovery groups, respectively) and for the majority of participants (6 out of 8 for each experiment; individual fits shown in Extended Data Fig. 6f, Extended Data Fig. 2c-e).

The COIN model explains memory recovery by creating a new memory only when existing memories cannot account for a perturbation, such as on the abrupt introduction of P^{+} and P^{-}, but not when a new perturbation is introduced gradually. This explains why deadaptation is slower following the removal of a gradually (vs. abruptly) introduced perturbation[17] (Extended Data Fig. 5p-s).

Memory updating depends on contextual inference

In the COIN model, contextual inference also controls how each existing memory is updated, that is *proper* learning (Fig. 1b, arrows 3). In the COIN model *all* memories are updated, with the updates scaled by their respective inferred responsibilities (Fig. $1\mathrm{f}_3$). This contrasts with models which only update a single memory[11,12] or update multiple memories independent of context[1,18]. To test this prediction, we examined the extent to which memories for two contexts were updated when we modulated their responsibilities by controlling the sensory cue and state feedback – the two observations that determine context responsibilities (Fig. 1b).

In many natural scenarios, sensory cues and state feedback provide consistent evidence about context (e.g. larger cups are heavier), and thus context responsibilities are approximately all-or-none (Fig. $1\mathrm{f}_3$). Thus to test for graded memory updating, we created conflicts between cues and state feedback (akin to a light, large cup). Specifically, participants experienced an extensive training phase designed to form separate memories for two contexts associated with a distinct cue (target location) and perturbation (Fig. 3a; context 1 = P^{+}_{1} and context 2 = P^{-}_{2}, with sub- and superscript specifying sensory cue and perturbation sign, respectively). These contexts switched randomly (with probability 0.5; Fig. 3b). As expected[19], participants formed separate memories for each context and expressed them appropriately based on the sensory cues (Extended Data Fig. 7a). In a subsequent test phase, we studied the updating of one of the memories, that associated with context 1, in response to exposure to a single trial of a potentially conflicting cue-feedback combination. To quantify single-trial learning for the memory associated with context 1, we assessed the adaptation of this memory using channel trials with the appropriate cue (cue 1) both before and after an exposure trial (Fig. 3c). The change in adaptation from the first to last channel trial of this 'triplet' (channel-exposure-channel) reflects single-trial learning in response to the exposure trial[4,5]. To bring adaptation back close to baseline before each triplet, we used sequences of washout trials, pairing P^0 with the sensory cues (P^0_1 and P^0_2).

The COIN model predicted that the responsibility of context 1, and hence the updating of the corresponding memory (as reflected in single-trial learning; Fig. 3d, column 2, Extended Data Fig. 4c), should exhibit a graded pattern that arises over training (Extended Data Fig. 7b): it should be greatest when the cue and state feedback on the exposure trial both provide evidence of context 1 (P^+_1 exposure trial), least when both provide evidence for context 2 (P^-_2 exposure trial) and intermediate when the two sources of evidence are in conflict (P^+_2 and P^-_1 exposure trials; see also Suppl. Inf. and Extended Data Fig. 7c-d for an analytical approximation). Comparing the two conditions with intermediate updating, due to the cues being paired with P^0 in the washout trials, we also expected the cue to have a weaker effect than the perturbation and therefore less updating of the memory for context 1 following exposure with P^-_1 than with P^+_2.

The pattern of single-trial learning in pre- and post-training confirmed the COIN model's qualitative predictions (Fig. 3d, column 1). Prior to training, there was no significant difference in single-trial learning across exposure conditions (two-way repeated-measures ANOVA, $F_{1,23} = 2.40$, $p = 0.135$ for cue, $F_{1,23} = 0.97$, $p = 0.335$ for perturbation). After learning, single-trial learning showed a gradation across conditions with a significant modulatory effect for both the cue and the perturbation ($F_{1,23} = 10.35$, $p = 3.82 \times 10^{-3}$ for cue, $F_{1,23} = 21.16$, $p = 1.26 \times 10^{-4}$ for perturbation, with no significant interaction, $F_{1,23} = 0.64$, $p = 0.432$; Extended Data Fig. 7e). The modulatory effects of the cue and the perturbation were not confined to separate subsets of participants (Fisher's exact test, odds ratio $= 1.0$, $p = 1.00$, see Methods and Extended Data Fig. 7f). After fitting to the data, the COIN model also accounted quantitatively for how single-trial learning changed during the training phase (Extended Data Fig. 7b). Taken together, the pattern of single-trial learning shows the gradation in memory updating (at an individual participant-level) predicted by the COIN model, with multiple memories updated in proportion to their responsibilities.

Apparent changes in learning rate

The COIN model also suggested an alternative account of classical results about apparent changes in learning rate under a variety of conditions. Fig. 4 shows three paradigms (column 1) with experimental data (column 2). What is common in all these cases is that the empirical finding of trial-to-trial changes in adaptation has been interpreted as *proper* learning, i.e. changes to existing memories (states). Thus differences between the magnitudes of these changes have been interpreted as differences in learning rate. For example, savings (Fig. 4a) refers to the phenomenon that learning the same perturbation a second time (even after washout) is faster than the first time[1,2,20,21]. In anterograde interference (Fig. 4b) learning a perturbation (P^-) is slower if an opposite perturbation (P^+) has been learned previously[3], with the amount of interference increasing with the length of experience of the first perturbation. The persistence of the environment has also been shown to affect single-trial learning (Fig. 4c)[4,5]: more consistent environments lead to increased levels of single-trial learning.

The COIN model suggests that changes in adaptation can occur without *proper* learning, simply through *apparent* learning, that is by changing the way existing memories are expressed (Fig. 1d-f, blue vs. cyan arrows). Therefore, apparent changes in learning rate in these paradigms may be due to changes in memory expression rather than changes in memory updating. To test this hypothesis, we simulated the COIN model using the parameters obtained by fitting each of the 40 participants in our experiments (Extended Data Fig. 3), thus providing parameter-free predictions. The COIN model reproduced the pattern of adaptation and single-trial learning seen in these paradigms (Fig. 4 and Extended Data Fig. 8, column 3; Extended Data Fig. 4d-f). Crucially, differences in the apparent learning rate were not driven by differences in either the proper learning rate (Kalman gain, see Methods) or the underlying state (column 4). Instead, they were driven by changes in contextual inference (column 5). For example, according to the COIN model, in savings P^{+} is expected with higher probability during the second exposure after having experienced it during the first exposure. Similarly, anterograde interference arises as more extended experience with P^{+} makes it less probable that a transition to other contexts (i.e. P^{-}) will occur. Finally, more (less) consistent environments lead to higher (lower) probabilities with which contexts are predicted to persist to the next trial, leading to more (less) memory expression, as reflected in single-trial learning. More generally, our analysis of the COIN model indicated that single-trial learning can be expressed mathematically as a mixture of two processes that both depend on contextual inference (see Suppl. Inf. and Extended Data Fig. 7c-d) and each of which can be dissected by the appropriate experimental manipulation: *proper* learning (as studied in Fig. 3) and *apparent* learning (as studied in Fig. 4c).

Cognitive mechanisms in contextual inference

In addition to providing a comprehensive account of the phenomenology of motor learning, the COIN model also suggests how specific cognitive mechanisms contribute to the underlying computations. For example, associating working memory with the maintenance of the currently estimated context probabilities explains how a working memory task can effectively lead to evoked recovery in a modified version of the spontaneous recovery paradigm[22] (see Suppl. Inf. and Extended Data Fig. 9a-d). Furthermore, identifying explicit and implicit forms of visuomotor learning with inferences in the model about state (i.e. estimate of visuomotor rotation) versus a bias parameter (i.e. sensory recalibration between the proprioceptive and visual locations of the hand), respectively, explains the complex time courses of these components of learning[23–25] (see Suppl. Inf. and Extended Data Fig. 9e-l).

Discussion

The COIN model puts the problem of learning a repertoire of memories — rather than a single motor memory — centre stage. Once this more general problem is considered, contextual inference becomes a key computation that unifies seemingly disparate data sets. By partitioning motor learning into two fundamentally different processes, contextual inference (Fig. 1b, top row) and state inference (Fig. 1b, bottom rows), the COIN model provides a principled framework for studying the neural bases of learning motor repertoires (see Suppl. Inf.).

In contrast to the COIN model, previous theories of motor learning typically did not have a notion of context[1,4,18]. In the few cases in which contextual motor learning was considered within a principled probabilistic framework[11,16,26], the generative models underlying learning did not incorporate fundamental properties of the environment (e.g. context transitions, cues or state dynamics) that are critical for explaining a number of learning phenomena. Consequently, previous models can only account for a subset of the data sets we model (Extended Data Table 1), which they were often hand-tailored to address.

There are deep analogies between the context-dependence of learning in the motor system and other learning systems, both in terms of their phenomenologies and the computational problems they are trying to solve[12,27–30]. However, there is one important conceptual issue that has been absent from work on contextual learning in other domains that our work has brought to the fore – the distinction between *proper* learning and *apparent* learning. We have shown that many features of motor learning arise not from the updating of existing memories (proper learning) but from changes in the extent to which existing memories are expressed (apparent learning). This distinction, and the role of contextual inference in both proper and apparent learning, is likely to be relevant to all forms of learning in which experience can be usefully broken down into discrete contexts – in the motor system and beyond.

References

1. Smith, M. A., Ghazizadeh, A. & Shadmehr, R. Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol.* **4**, e179 (2006).
2. Kitago, T., Ryan, S., Mazzoni, P., Krakauer, J. W. & Haith, A. M. Unlearning versus savings in visuomotor adaptation: comparing effects of washout, passage of time, and removal of errors on motor memory. *Front. Hum. Neurosci.* **7**, 307 (2013).
3. Sing, G. C. & Smith, M. A. Reduction in learning rates associated with anterograde interference results from interactions between different timescales in motor adaptation. *PLoS Comput. Biol.* **6**, e1000893 (2010).
4. Herzfeld, D. J., Vaswani, P. A., Marko, M. K. & Shadmehr, R. A memory of errors in sensorimotor learning. *Science* **345**, 1349–1353 (2014).
5. Gonzalez Castro, L. N., Hadjiosif, A. M., Hemphill, M. A. & Smith, M. A. Environmental consistency determines the rate of motor adaptation. *Curr. Biol.* **24**, 1050–1061 (2014).
6. Mcdougle, S. D. *et al.* Credit assignment in movement-dependent reinforcement learning. *Proc. Natl. Acad. Sci.* **113**, 6797–6802 (2016).
7. Donchin, O., Francis, J. T. & Shadmehr, R. Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: theory and experiments in human motor control. *J. Neurosci.* **23**, 9032–9045 (2003).
8. Thoroughman, K. A. & Shadmehr, R. Learning of action through adaptive combination of motor primitives. *Nature* **407**, 742–747 (2000).
9. Shadmehr, R., Smith, M. A. & Krakauer, J. W. Error correction, sensory prediction, and adaptation in motor control. *Ann. Rev. Neurosci.* **33**, 89–108 (2010).
10. Wolpert, D. M. & Kawato, M. Multiple paired forward and inverse models for motor control. *Neural Netw.* **11**, 1317–1329 (1998).
11. Oh, Y. & Schweighofer, N. Minimizing precision-weighted sensory prediction errors via memory formation and switching in motor adaptation. *J. Neurosci.* 9237–9250 (2019).
12. Gershman, S. J., Radulescu, A., Norman, K. A. & Niv, Y. Statistical computations underlying the dynamics of memory updating. *PLoS Comput. Biol.* **10**, e1003939 (2014).
13. Hulst, T. *et al.* Cerebellar degeneration reduces memory resilience after extended training. *bioRxiv* (2020).
14. Pekny, S. E., Criscimagna-Hemminger, S. E. & Shadmehr, R. Protection and expression of human motor memories. *J. Neurosci.* **31**, 13829–13839 (2011).
15. Rescorla, R. A. & Heth, C. D. Reinstatement of fear to an extinguished conditioned stimulus. *J. Exp. Psychol.: Animal Behavior Processes* **1**, 88–96 (1975).
16. Berniker, M. & Körding, K. Estimating the sources of motor errors for adaptation and generalization. *Nat. Neurosci.* **11**, 1454–1461 (2008).
17. Taylor, J. A., Wojaczynski, G. J. & Ivry, R. B. Trial-by-trial analysis of intermanual transfer during visuomotor adaptation. *J. Neurophysiol.* **106**, 3157–3172 (2011).
18. Körding, K. P., Tenenbaum, J. B. & Shadmehr, R. The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nat. Neurosci.* **10**, 779–786 (2007).
19. Heald, J. B., Ingram, J. N., Flanagan, J. R. & Wolpert, D. M. Multiple motor memories are learned to control different points on a tool. *Nat. Hum. Behav.* **2**, 300–311 (2018).
20. Coltman, S. K., Cashaback, J. G. A. & Gribble, P. L. Both fast and slow learning processes contribute to savings following sensorimotor adaptation. *J. Neurophysiol.* **121**, 1575–1583 (2019).
21. Huang, V. S., Haith, A., Mazzoni, P. & Krakauer, J. W. Rethinking motor learning and savings in adaptation paradigms: model-free memory for successful actions combines with internal models. *Neuron* **70**, 787–801 (2011).
22. Keisler, A. & Shadmehr, R. A shared resource between declarative memory and motor memory. *J. Neurosci.* **30**, 14817–14823 (2010).

23. Mcdougle, S. D., Ivry, R. B. & Taylor, J. A. Taking aim at the cognitive side of learning in sensorimotor adaptation tasks. *Trends Cogn. Sci.* **20**, 535–544 (2016).
24. McDougle, S. D., Bond, K. M. & Taylor, J. A. Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *J. Neurosci.* **35**, 9568–9579 (2015).
25. Miyamoto, Y. R., Wang, S. & Smith, M. A. Implicit adaptation compensates for erratic explicit strategy in human motor learning. *Nat. Neurosci.* **23**, 443–455 (2020).
26. Haruno, M., Wolpert, D. M. & Kawato, M. MOSAIC model for sensorimotor learning and control. *Neural Comput.* **13**, 2201–2220 (2001).
27. Gershman, S. J., Blei, D. M. & Niv, Y. Context, learning, and extinction. *Psychol. Rev.* **117**, 197–209 (2010).
28. Sanders, H., Wilson, M. A. & Gershman, S. J. Hippocampal remapping as hidden state inference. *eLife* **9**, e51140 (2020).
29. Collins, A. & Koechlin, E. Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol.* **10**, e1001293 (2012).
30. Collins, A. G. E. & Frank, M. J. Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol. Rev.* **120**, 190–229 (2013).

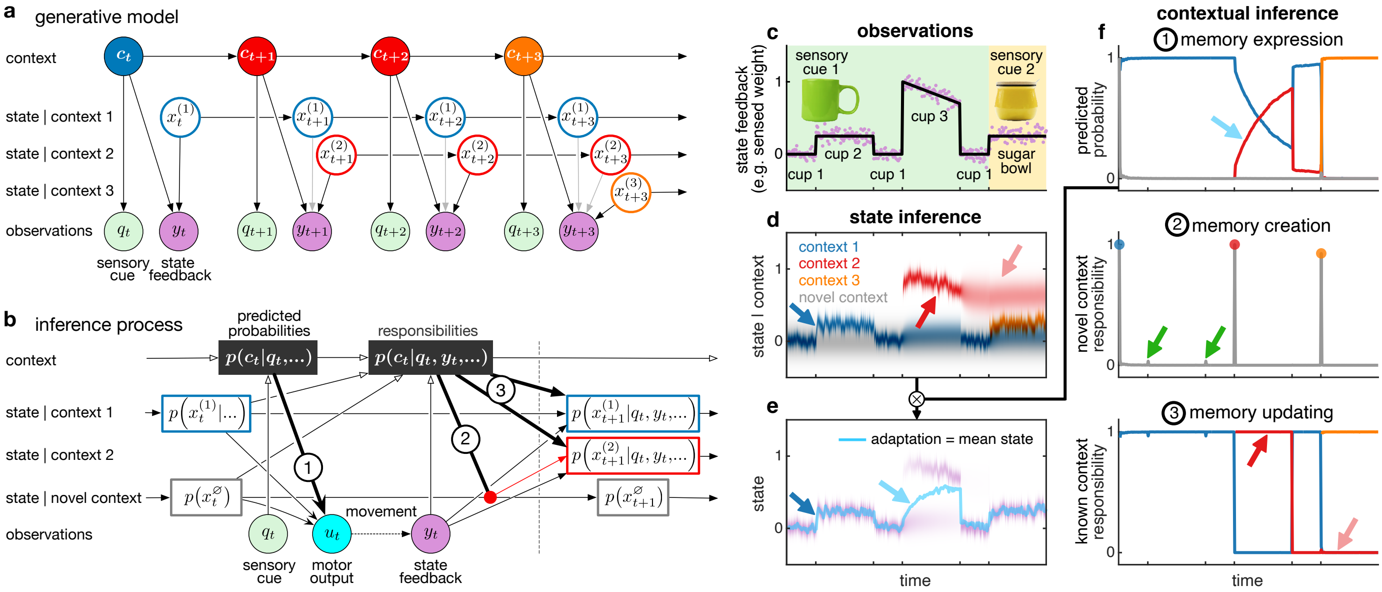


Fig. 1 | Contributions of contextual inference to motor learning in the COIN model. a, Generative model. A (potentially) infinite number of discrete contexts c_t (colours) exist that transition as a Markov process. Each context j is associated with a time-varying state $x_t^{(j)}$. The active context can generate a sensory cue q_t independent of movement (e.g. the visual appearance of an object) and also determines which state is observed (with noise) as state feedback y_t as a consequence of movement (e.g. object weight, black vs. grey arrows). **b,** Inference process. The learner infers contexts and states (and parameters, not shown) based on observed sensory cues and state feedback. Before movement, predicted context probabilities $p(c_t \mid q_t, \ldots)$ are computed by fusing prior expectations from the previous time point (where ... refers to all observations before time t) with the likelihood of the current sensory cue q_t. For each known context, a predicted distribution over its current state $p(x_t^{(j)} \mid \ldots)$ is represented. A potential novel context is always represented, with a stationary state distribution $p(x_t^{\varnothing})$. Motor output u_t is the average of the states of the known and novel contexts, weighted by their predicted probabilities (arrow 1). Movement results in state feedback y_t, which updates the predicted context probabilities to context responsibilities $p(c_t \mid q_t, y_t, \ldots)$. A new memory is instantiated with a probability that is the responsibility of the novel context (arrow 2, showing the creation of a red context, initialised as a copy of the state distribution of the novel context). Responsibilities also determine the degree to which state feedback is used to update the predicted state distribution $p(x_{t+1}^{(j)} \mid q_t, y_t, \ldots)$ of each context (arrows 3). **c,** Simulated time series of sensory cues (background colour for object appearance) and state feedback observations (noisy weight, purple) when handling visually-identical cups and a sugar bowl of varying weights (black line, arbitrary scale). The weight of cup 3 decreases as liquid is poured from it, other objects have constant weights. **(d-f)** The COIN model applied to the observations in **c**. **d,** Predicted state distributions for the three contexts inferred by the model and a novel context. **e,** The predicted state distribution (purple) is a mixture of the individual contexts' predicted state distributions (**d**) weighted by their predicted probabilities (**f**$_1$). The motor output (adaptation, cyan line) is the mean of the predicted state distribution. Intensity of colours in **d** and purple in **e** indicates probability density, linearly scaled between 0 and the maximum of the corresponding density. **f,** Contextual inferences (colours as in **d**). **1.** Predicted probability (before state feedback) of each known context and a novel context. **2.** Responsibility (context probability after state feedback) of a novel context. Coloured circles show memory creation events. The novel context responsibility is insufficient to generate a new memory when transitioning to and from cup 2 (green arrows). **3.** Responsibility of each known context. See text for arrow explanations in **d-f**.

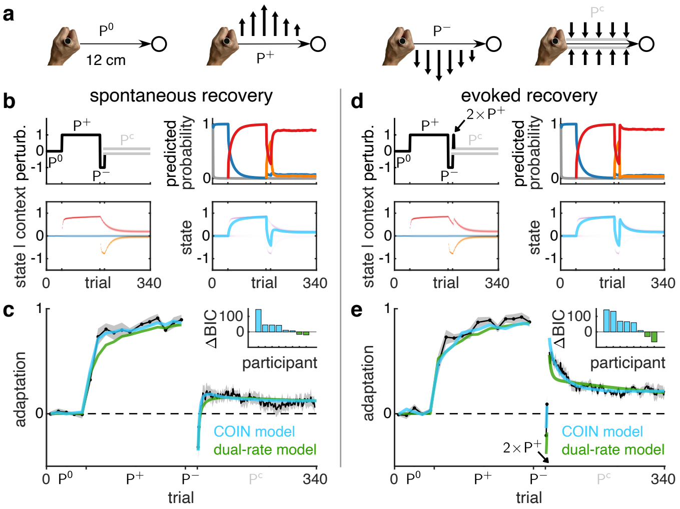


Fig. 2 | Memory creation and expression accounts for spontaneous and evoked recovery. a, Participants made reaching movements (thin horizontal arrows) to a target (circle) while holding the handle of a robotic manipulandum that could generate forces (thick vertical arrows). For clarity, schematic not to scale. The manipulandum could either be passive (null field, P^0) or generate a velocity-dependent force field that acted to the left (P^+) or right (P^-) of the current movement direction. Channel trials (P^c) were used to assess adaptation by constraining the hand to a straight channel (grey lines) to the target and measuring the forces generated by the participant into the virtual channel walls. **b,** Simulation of the spontaneous recovery paradigm with the COIN model (parameters fit to average data in **c** & **e** simultaneously). Top left: perturbation (black) and channel-trial phase (grey). Bottom left: predicted state distributions of inferred contexts as in Fig. 1d (for clarity we omit the novel context here and in subsequent figures). Top right: predicted probability of contexts as in Fig. 1d. Bottom right: predicted state distribution (purple) and its mean (cyan) as in Fig. 1e. Note that full state distributions are inferred in bottom left and right but they appear narrow due to fitting to the average of all participants' data (see Methods). **c,** Mean adaptation (black, ± SEM across n = 8 participants) on the channel trials of the spontaneous recovery paradigm. The cyan and green lines show model fits (mean of individual participant fits) of the COIN (7 parameters) and dual-rate models (5 parameters), respectively. Inset shows ΔBIC (nats) for individual participants, positive favours the COIN model. **d-e**, As in **b-c** for the evoked recovery paradigm (n = 8) in which the 3rd and 4th trials in the channel-trial phase were replaced by P^+ trials (black arrow). For COIN model parameters see Extended Data Fig. 3.

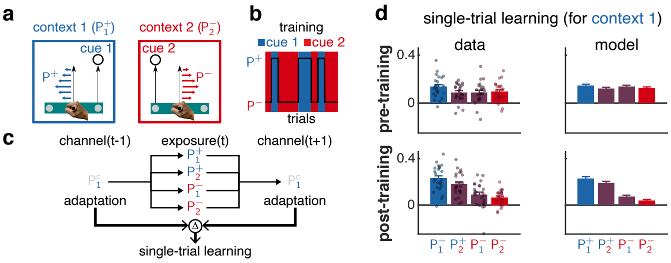


Fig. 3 | Memory updating depends on contextual inference. a, Participants experienced two contexts defined by a sensory cue (right or left target) paired with a perturbation sign (P^+ or P^-). Participants moved a control point (right vs. left , grey disk) on a virtual bar to the corresponding target[19]. For clarity, schematic not to scale. The colours of cues and perturbations indicate the context to which they are associated (blue and red for context 1 and 2, respectively). **b,** Training: cues (background colour) were consistently paired with perturbations (black line) randomly selected on each trial (only a few trials shown for clarity). **c,** Triplets: two channel trials (both with cue 1, P_1^c) bracket an 'exposure' trial that uses one of the four possible cue-perturbation combinations. Single-trial learning for the memory associated with context 1 is measured as the difference (Δ) in adaptation across the two channel trials. **d** Single-trial learning for context 1 before (top) and after (bottom) training. Experimental data (mean ± SEM, column 1) across n = 24 participants (dots). Positive values indicate single-trial learning consistent with the exposure trial perturbation (increase following P^+ and decrease following P^-). The average (± SEM across participants, column 2) of the individual COIN model fits (8 parameters each, Extended Data Fig. 3).

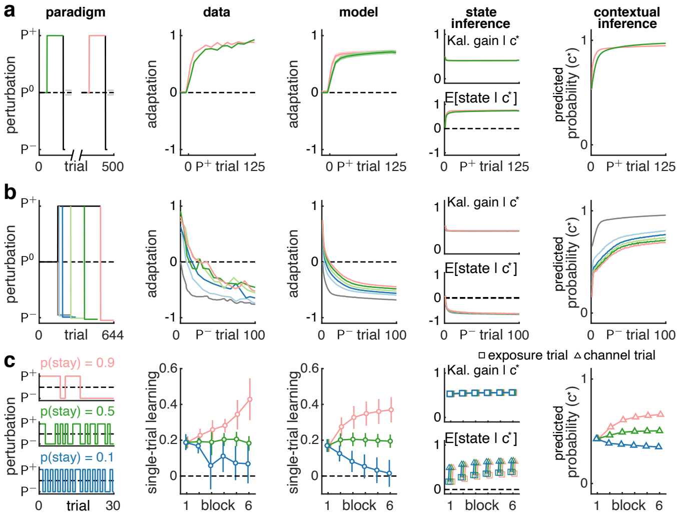


Fig. 4 | Contextual inference underlies apparent changes in learning rate. The COIN model applied to three phenomena: savings (**a**), anterograde interference (**b**) and the effect of environmental consistency on single-trial learning (**c**). Column 1: experimental paradigms (lines as in previous figures, colours highlight key comparisons). Note the lines showing P$^-$ perturbations in **b** have been separated vertically for clarity. Column 2: experimental data replotted from Ref. 20 (**a**), Ref. 3 (**b**) and Ref. 4 (**c**). Column 3: output of COIN model averaged over 40 parameter sets obtained from fits to individual participants in the experiments shown in Figs. 2 and 3 (7 parameters, Extended Data Fig. 3). Error bars show SEM based on the number of participants in the original experiments (n = 46 in **a**, n = 50 in **b** and n = 27 in **c**). Columns 4-5: COIN model inferences with regard to the context (c^*) that is most relevant to the perturbation to which adaptation is measured. Specifically, c^* is the context with the highest responsibility on the given trial (that associated with P$^+$ in **a** and P$^-$ in **b**) or, as in Fig. 3d (also single-trial learning), the context with the highest predicted probability on the second channel trial of a triplet (that associated with P$^+$, **c**). Column 4: Kalman gain (top) and mean of the predicted state distribution (bottom) for the relevant context c^*. Column 5: Predicted probability of the relevant context c^*. Grey lines in **b** represent initial adaptation to P$^+$ and have been sign inverted in columns 2-3 and the bottom panel of column 4. Data in **c** shows averages within blocks, with the bottom panel in column 4 showing separate averages for exposure (squares) and subsequent channel trials (triangles).

Methods

Here, we provide an overview of the methods. For full details see Suppl. Inf.

Participants

Forty right-handed, neurologically-healthy participants (18 males and 22 females; age 27.7 ± 5.6 yr, mean ± s.d.) participated in two experiments, which had been approved by the Cambridge Psychology Research Ethics Committee and the Columbia University IRB (AAAR9148). All participants provided written informed consent.

Experimental apparatus

Experiments were performed using a vBOT planar robotic manipulandum with virtual-reality system and air table[31]. Participants grasped the handle of the manipulandum with their right hand while their forearm was supported on an air sled and moved their hand in the horizontal plane.

The manipulandum controlled a virtual "object" that was displayed centred on the hand and translated with hand movements as participants made repeated movements from a home position to a target located 12 cm distally in the sagittal direction.

On each trial, the vBOT could either generate no forces (P^0, null field), a velocity-dependent curl force field (P^+ or P^- perturbation depending on the direction of the field) or a force channel (P^c, channel trials). For the curl force field, the force generated on the hand was given by

$$\begin{bmatrix} F_\mathrm{x} \\ F_\mathrm{y} \end{bmatrix} = g \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \tag{1}$$

where F_x, F_y, $\dot{x}$ and $\dot{y}$ are the forces and velocities at the handle in the x (transverse) and y (sagittal) directions respectively. The gain was set to ±15 $\mathrm{N{\cdot}s{\cdot}m^{-1}}$, with the sign specifying the direction of the curl field (counterclockwise or clockwise, which were assigned to P^+ and P^-, counterbalanced across participants). On channel trials, the hand was constrained to move along a straight line to the target by simulating channel walls on each side of the straight line as stiff springs (3,000 $\mathrm{N{\cdot}m^{-1}}$) with damping (140 $\mathrm{N{\cdot}s{\cdot}m^{-1}}$)[32,33].

Experiment 1: spontaneous and evoked recovery

Sixteen participants were assigned to either a spontaneous (n = 8) or evoked (n = 8) recovery group. The virtual object controlled by participants was simply a cursor.

Participants in the spontaneous recovery group performed a version of the standard spontaneous recovery paradigm[1]. A pre-exposure phase (50 trials) with a null field (P^0) was followed by an exposure phase (125 trials) with P^+. Participants then underwent a counter-exposure phase of 15 trials with the opposite perturbation (P^-). This was followed by a channel-trial phase (150 channel trials, P^c). In the pre-exposure and exposure phases, to assess adaptation, each block of 10 trials had one channel trial (P^c) in a random location (not the first). A 45 s rest break was given after trial 60 of the exposure phase, followed by an additional 5 P^+ trials prepended to the next block.

The evoked recovery group experienced the identical paradigm to the spontaneous recovery group except that the 3rd and 4th trials of the channel-trial phase were replaced with P^{+} trials (Fig. 2d).

Experiment 2: memory updating

Twenty-four participants performed the memory updating experiment. The paradigm is based on the control point experiment described in Ref. 19 in which perturbations P^0_1, P^0_2, P^+_1, P^+_2, P^-_1 and P^-_2 are presented with one of two possible sensory cues (different control points on a rectangular virtual object, denoted by subscripts). The experiment consisted of a pre-training, training and post-training phase. In the pre-training and post-training phases, participants performed blocks of trials consisting of a variable number (8, 10 or 12 in the pre-training phase and 2, 4 or 6 in the post-training phase) of washout trials (an equal number of P^0_1 and P^0_2 in a pseudorandom order) followed by 1 of 4 possible 'triplets'. Each triplet consisted of 2 channel trials (both with cue 1, P^c_1) bracketing a cue-perturbation 'exposure' trial (P^+_1, P^+_2, P^-_1 or P^-_2, see main text and Fig. 3c). Each of the 4 triplet types was experienced once every 4 blocks, using pseudorandom permutations, with a total of 16 blocks in the pre-training phase and 32 blocks in the post-training phase.

In the training phase (Fig. 3b), participants performed 24 blocks each consisting of 62–70 trials. The key feature of each block was that 32 force-field trials (equal number of P^+_1 and P^-_2 in a pseudorandom order) was followed by 2 triplets (with exposure trials of P^+_1 and P^-_2). Each triplet was preceded by a variable number of washout trials (equal number of P^0_1 and P^0_2 in a pseudorandom order) to bring adaptation back close to baseline. For full details of the block structure see Suppl. Inf.

The control point assigned to sensory cue 1 (used on all triplet channel trials) and sensory cue 2 was counterbalanced across participants as was the direction of force field assigned to P^+ and P^-.

Data analysis

On each channel trial, we linearly regressed the time series of actual forces generated by participants into the channel wall against the ideal forces that would fully compensate for the forces on a force-field trial[1]. The offset of the regression was constrained to zero, and we used the slope as our (dimensionless) measure of adaptation.

To identify changes in single-trial learning between triplets in the memory updating experiment, two-way repeated-measures ANOVAs were performed with factors of cue (2 levels: cue 1 and cue 2) and perturbation (2 levels: P^+ and P^-). To test whether the modulatory effects of cue and perturbation were confined to separate subsets of participants, we quantified the effect of each by computing, on an individual-participant basis, the following contrasts in single-trial learning: $\mathrm{P}^+_1 + \mathrm{P}^-_1 - \mathrm{P}^+_2 - \mathrm{P}^-_2$ (cue effect) and $\mathrm{P}^+_1 + \mathrm{P}^+_2 - \mathrm{P}^-_1 - \mathrm{P}^-_2$ (perturbation effect). We then split participants into 2×2 groups based on whether each effect was below or above the median of each effect and performed a Fisher's exact test on the resulting 2×2 histogram (see Suppl. Inf. for details).

All statistical tests were two-sided with significance set to $p < 0.05$. Data analysis was performed using MATLAB R2020a.

COIN generative model

Fig. 1a shows the graphical model for the generative model. At each time step $t = 1, \ldots, T$ there is a discrete latent variable (the context) $c_t \in \{1, \ldots, \infty\}$ that evolves as a Markov process:

$$c_t \mid c_{t-1}, \mathbf{\Pi} \sim \text{Discrete}\big(\boldsymbol{\pi}_{c_{t-1}}\big)\,, \tag{2}$$

where $\mathbf{\Pi} = (\boldsymbol{\pi}_j)_{j=1}^{\infty}$ is the transition probability matrix and $\boldsymbol{\pi}_j = (\pi_{jk})_{k=1}^{\infty}$ is its j^{th} row containing the transition probabilities from context j to each context k (including itself). In principle, there are an infinite number of rows and columns in this matrix. However, in practice, generation and inference can both be accomplished using finite-sized matrices by placing a nonparametric prior on the matrix (see below).

Each context j is associated with a continuous (scalar) latent variable $x_t^{(j)}$ (the state, e.g. the strength of a force field) that evolves according to its own linear-Gaussian dynamics independently of all other states:

$$x_t^{(j)} = a^{(j)}\, x_{t-1}^{(j)} + d^{(j)} + w_t^{(j)} \qquad w_t^{(j)} \sim \mathcal{N}\big(0, \sigma_{\text{q}}^2\big)\,, \tag{3}$$

where $a^{(j)}$ and $d^{(j)}$ are the context-specific state retention factor and drift, respectively, and σ_{q}^2 is the variance of the process noise (shared across contexts). Each state is assumed to have existed for long enough that its prior for the first time it is observed is its stationary distribution:

$$\lim_{t\to\infty} x_t^{(j)} \sim \mathcal{N}(d^{(j)}/(1-a^{(j)}), \sigma_{\text{q}}^2/(1-[a^{(j)}]^2)). \tag{4}$$

At each time step, a continuous (scalar) observation y_t (the state feedback) is emitted from the state associated with the current context:

$$y_t = x_t^{(c_t)} + v_t \qquad v_t \sim \mathcal{N}\big(0, \sigma_{\text{r}}^2\big)\,, \tag{5}$$

where σ_{r}^2 is the variance of the observation noise (also shared across contexts).

In addition to the state feedback, a discrete observation (the sensory cue) $q_t \in \{1, \ldots, \infty\}$ is also emitted. The distribution of sensory cues depends on the current context:

$$q_t \mid c_t, \mathbf{\Phi} \sim \text{Discrete}\big(\boldsymbol{\phi}_{c_t}\big)\,, \tag{6}$$

where $\mathbf{\Phi} = \big(\boldsymbol{\phi}_j\big)_{j=1}^{\infty}$ is the cue probability matrix (which, in principle, is also doubly infinite in size but can be treated as finite in practice) and $\boldsymbol{\phi}_j = (\phi_{jk})_{k=1}^{\infty}$ is its j^{th} row containing the probability of each cue k in context j.

In order to make this infinite-dimensional switching state-space model well-defined, we place hierarchical Dirichlet process priors[34] on the transition and cue probability matrices. The transition probability matrix is generated in two steps (Extended Data Fig. 1a). First, an infinite set of global probabilities for transitioning into each context $\boldsymbol{\beta} = (\beta_j)_{j=1}^{\infty}$ ('global transition probabilities') is generated by sampling from a GEM (Griffiths, Engen and McCloskey) distribution:

$$\boldsymbol{\beta} \mid \gamma \sim \text{GEM}(\gamma)\,, \tag{7}$$

where $0 \le \beta_j \le 1$ and $\sum_{j=1}^{\infty} \beta_j = 1$, as required for a set of probabilities. The global transition probabilities decay exponentially as a function of j in expectation, with the hyperparameter γ controlling the rate of decay and thus the effective number of contexts: large γ implies a large number of small-probability contexts (slow decay from a relatively small initial probability), whereas small γ implies a smaller number of relatively large-probability contexts (fast decay from a relatively large initial probability).

Second, for each context (row of the transition probability matrix), an infinite set of local (context-specific) probabilities for transitioning into each context $\boldsymbol{\pi}_j = (\pi_{jk})_{k=1}^{\infty}$ ('local transition probabilities') are generated via a 'sticky' variant[35] of the Dirichlet process (DP):

$$\boldsymbol{\pi}_j \mid \alpha, \boldsymbol{\beta}, \kappa \sim \mathrm{DP}\left(\alpha + \kappa, \frac{\alpha\,\boldsymbol{\beta} + \kappa\,\boldsymbol{\delta}_j}{\alpha + \kappa}\right), \tag{8}$$

where $0 \leq \pi_{jk} \leq 1$ and $\sum_{k=1}^{\infty} \pi_{jk} = 1$, as required for a set of probabilities, and $\boldsymbol{\delta}_j$ is an infinite-dimensional one-hot vector with the j^{th} element set to 1 and all other elements set to 0. The mean (base) distribution of the Dirichlet process is $(\alpha\,\boldsymbol{\beta} + \kappa\,\boldsymbol{\delta}_j)/(\alpha + \kappa)$, with large $\alpha + \kappa$ reducing variability around this mean (for a tutorial on the Dirichlet process see Ref. 36). Thus the concentration parameter α controls the resemblance of local transition probabilities to the global transition probabilities $\boldsymbol{\beta}$. The self-transition bias parameter $\kappa > 0$ controls the resemblance of local transition probabilities to $\boldsymbol{\delta}_j$ (i.e. a certain self-transition, $c_t = c_{t-1} = j$). This self-transition bias expresses the fact that a context often persists for several time steps before switching (i.e. that contexts are 'sticky'), such as when an object is manipulated for an extended period of time.

Note that the rows of the transition probability matrix are dependent as their expected values (the base distributions of the corresponding Dirichlet processes) contain a shared term, the global transition distribution $\boldsymbol{\beta}$. This dependency, controlled by α, captures the intuitive notion that contexts that are common in general (i.e. have a large global transition probability) will be transitioned to frequently from all contexts.

The cue probability matrix $\boldsymbol{\Phi} = (\boldsymbol{\phi}_j)_{j=1}^{\infty}$ is generated using an analogous (non-sticky) hierarchical construction:

$$\boldsymbol{\beta}^{\mathrm{e}} \mid \gamma^{\mathrm{e}} \sim \mathrm{GEM}(\gamma^{\mathrm{e}}) \qquad \boldsymbol{\phi}_j \mid \alpha^{\mathrm{e}}, \boldsymbol{\beta}^{\mathrm{e}} \sim \mathrm{DP}(\alpha^{\mathrm{e}}, \boldsymbol{\beta}^{\mathrm{e}}), \tag{9}$$

where γ^{e} determines the distribution of the global cue probabilities $\boldsymbol{\beta}^{\mathrm{e}}$, and α^{e} determines the across-context variability of local cue probabilities around the global cue probabilities.

In order to allow full Bayesian inference over the parameters governing the state dynamics $\boldsymbol{\omega}^{(j)} = \left[a^{(j)}\ d^{(j)}\right]^{\mathsf{T}}$, we also place a prior on these parameters. For this, we use a bivariate normal distribution (truncated for $a^{(j)}$ between 0 and 1):

$$\boldsymbol{\omega}^{(j)} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{TN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{10}$$

where $\boldsymbol{\mu} = \left[\mu_{\mathrm{a}}\ 0\right]^{\mathsf{T}}$ and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_{\mathrm{a}}^2, \sigma_{\mathrm{d}}^2)$ is a diagonal covariance matrix. Here we have set the prior mean of $d^{(j)}$ to zero under the assumption that positive and negative drifts are equally probable.

Inference in the COIN model

At each time step $t = 1, \ldots, T$, the goal of inference is to compute the joint posterior distribution $p(\boldsymbol{\Theta}_t \mid y_{1:\tau}, q_{1:\tau'})$ of all quantities $\boldsymbol{\Theta}_t = \{c_t, \{x_t^{(j)}, \boldsymbol{\omega}^{(j)}\}_{j=1}^{\infty}, \boldsymbol{\beta}, \boldsymbol{\Pi}, \boldsymbol{\beta}^{\mathrm{e}}, \boldsymbol{\Phi}\}$ that are not directly observed by the learner: the current context c_t, the current state of each context $x_t^{(j)}$, the parameters governing the state dynamics in each context $\boldsymbol{\omega}^{(j)}$, the context transition parameters (global $\boldsymbol{\beta}$ and local $\boldsymbol{\Pi}$ transition probabilities) and the cue emission parameters (global $\boldsymbol{\beta}^{\mathrm{e}}$ and local $\boldsymbol{\Phi}$ cue probabilities) based on the sequence of state feedback $y_{1:\tau}$ and sensory cue observations $q_{1:\tau'}$ made until time τ and τ', respectively (with τ and τ' each being either t or $t-1$, see below). In principle, this posterior is fully determined by the generative model defined in the previous section and can be obtained in a sequential manner by recursively propagating ('filtering') the joint posterior from one time point to the next after each new set of observations is made. As exact inference is infeasible, we use a sequential Monte Carlo method known as particle learning that computes an approximation to this filtered posterior[37,38]. We extensively validated the accuracy of this method (Extended Data Fig. 2a-b). The details of the inference method are given in Suppl. Inf. Here we only describe how the approximate posterior is used to obtain the main model-derived quantities plotted in the paper.

The predicted probability of context $j \in \{1, \ldots, J, \varnothing\}$, where J is the number of known contexts and $\varnothing$ is the novel context, on trial t (computed after observing the cue but before observing the state feedback; Fig. 1f$_1$ and corresponding panels in later figures) is

$$p(c_t = j \mid q_t, \ldots) = \int p(c_t = j, \boldsymbol{\Theta}_t \backslash c_t \mid q_t, \ldots) \mathrm{d}\boldsymbol{\Theta}_t \backslash c_t, \tag{11}$$

where $\boldsymbol{\Theta}_t \backslash c_t$ denotes the set $\boldsymbol{\Theta}_t$ excluding c_t and ... represents all observations before time t (as in Fig. 1). The responsibility of context j on trial t (computed after observing both the cue and the state feedback; Fig. 1f$_{2-3}$ and corresponding panels in later figures) is

$$p(c_t = j \mid q_t, y_t, \ldots) = \int p(c_t = j, \boldsymbol{\Theta}_t \backslash c_t \mid q_t, y_t, \ldots) \mathrm{d}\boldsymbol{\Theta}_t \backslash c_t. \tag{12}$$

The predicted state distribution for context j on trial t (computed before observing the state feedback; Fig. 1d and corresponding panels in later figures) is

$$p(x_t^{(j)} \mid \ldots) = \int p(x_t^{(j)}, \boldsymbol{\Theta}_t \backslash x_t^{(j)} \mid \ldots) \mathrm{d}\boldsymbol{\Theta}_t \backslash x_t^{(j)}, \tag{13}$$

where $\boldsymbol{\Theta}_t \backslash x_t^{(j)}$ denotes the set $\boldsymbol{\Theta}_t$ excluding $x_t^{(j)}$. The mean of this distribution $\hat{x}_t^{(j)}$ can be shown to evolve across trials (see Suppl. Inf.) as

$$\hat{x}_{t+1}^{(j)} = \mathbb{E}_{p(a^{(j)} \mid c_t, q_t, y_t, \ldots)}[a^{(j)}] \left(\hat{x}_t^{(j)} + p(c_t = j \mid q_t, y_t, \ldots)\, k_t^{(j)}\, e_t^{(j)} \right) + \mathbb{E}_{p(d^{(j)} \mid c_t, q_t, y_t, \ldots)}[d^{(j)}], \tag{14}$$

where $\mathbb{E}_{p(a^{(j)} \mid c_t, q_t, y_t, \ldots)}[a^{(j)}]$ denotes the expected value of $a^{(j)}$ with respect to the distribution $p(a^{(j)} \mid c_t, q_t, y_t, \ldots)$, $e_t^{(j)} = y_t - \hat{x}_t^{(j)}$ is the prediction error for context j and $k_t^{(j)}$ corresponds to the 'Kalman gain' for context j, which we plot in Fig. 4. Note that this update is scaled by the context's responsibility $p(c_t = j \mid q_t, y_t, \ldots)$, which underlies the effect of contextual inference on memory updating (arrows 3 in Fig. 1b).

The 'overall' predicted state distribution on trial t (i.e. the predicted state distribution of the context that is currently active, and of which the identity the learner cannot know with certainty; purple distribution in Fig. 1e and corresponding panels in later figures) is computed by integrating out the context from Eq. 13 using the predicted probabilities from Eq. 11 (arrow 1 in Fig. 1b):

$$\mathbb{E}_{p(c_t \mid q_t, \ldots)}[p(x_t^{(c_t)} \mid \ldots)] = \sum_{j = \{1, \ldots, J, \varnothing\}} p(x_t^{(j)} \mid \ldots)\, p(c_t = j \mid q_t, \ldots). \tag{15}$$

The motor output u_t of the learner (cyan line in Fig. 1e and corresponding panels in later figures) is the mean of this predicted state distribution:

$$u_t = \sum_{j = \{1, \ldots, J, \varnothing\}} \hat{x}_t^{(j)}\, p(c_t = j \mid q_t, \ldots). \tag{16}$$

Applying the COIN model to experimental data

Applying the COIN model to experimental data required solving two additional challenges. First, participants' state feedback observations are *hidden* from the perspective of the experimenter, as they are noisy realisations of the ‘true’ underlying states (Eq. 5). To appropriately account for our uncertainty about the state feedback participants actually observed, we computed the *distribution* of COIN model inferences by integrating over the possible sequences of state feedback observations $y_{1:T}$ given the sequence of true states (experimentally-applied perturbations) $x^*_{1:T}$[39]. Specifically, on each trial, x^*_t was assigned a value of 0 (null-field trials), $+1$ (P^+ perturbation trials) or -1 (P^- perturbation trials) and y_t was assumed to be distributed around x^*_t with i.i.d. zero-mean Gaussian observation noise of variance σ^2_r (Eq. 5), except on channel trials (P^c) where we treated y_t as unobserved, as the state (the magnitude of a force field) was not observed by the participants on those trials. Note that the distribution of state feedback given the true state $p(y_t|x^*_t)$ shares the same parameters as those underlying the COIN model inferences as it is self-consistently defined by the generative model. All figures showing COIN model inferences applied to experimental data (i.e. all but Fig. 1) show the quantities described in the previous section after the state feedback has been integrated out (Fig. 1d-f shows COIN model inferences conditioned on the state feedback sequence shown in Fig. 1c).

Second, real participants’ behaviour can always be subject to influences not explicitly included in the COIN model. In order to account for these uncontrolled and unmodelled factors, we introduced a phenomenological ‘motor noise’ component that related the motor output u_t of the COIN model (Eq. 16) to the experimentally measured adaptation a_t via i.i.d. zero-mean Gaussian noise:

$$a_t \sim \mathcal{N}\big(u_t, \sigma^2_\mathrm{m}\big)\,, \tag{17}$$

where σ_m is the standard deviation of the motor noise.

Model fitting and model comparison

In Experiments 1 and 2, we fit the parameters of the COIN model ϑ to participants’ data by maximising the data log likelihood using Bayesian adaptive direct search (BADS)[40]. In Experiment 1, $\vartheta = \{\sigma_\mathrm{q}, \mu_\mathrm{a}, \sigma_\mathrm{a}, \sigma_\mathrm{d}, \alpha, \rho, \sigma_\mathrm{m}\}$, where

$$\rho = \kappa/(\alpha + \kappa) \tag{18}$$

is the normalised self-transition bias parameter. In Experiment 2, which included sensory cues, an additional parameter α^e was also fit. In Experiment 1, we also fit a two-state (dual-rate) and three-state state-space model to the data of individual participants by minimising the mean squared error using MATLAB’s fmincon and BADS. In all cases, optimisation was performed from 30 random initial parameter settings (see Suppl. Inf.).

To perform model comparison for individual participants, we calculated the Bayesian information criterion (BIC). A BIC difference of greater than 4.6 nats (a Bayes factor of greater than 10) is considered to provide strong evidence in favour of the model with the lower BIC value[41]. To perform model comparison at the group level, we calculated the group-level BIC, which is the sum of BICs over individuals[42].

Parameter and model recovery

We used the parameters from the fits of the COIN and dual-rate models to the data of each participant in the spontaneous and evoked recovery experiments to generate 10 synthetic data sets per model class (COIN and dual-rate) for each participant from the corresponding experiment. In the dual-rate model, the only source of variability across the different synthetic data sets for a given participant was motor noise. In contrast, for the COIN model, sensory noise provided another source of variability in addition to motor noise. We fit both model classes to each synthetic data set as we did with real data (see above).

For parameter recovery (Extended Data Fig. 2c), we compared the COIN model parameters that were used to generate the synthetic data (‘true’ parameters) with the COIN model parameters fit to these synthetic data sets (‘recovered’ parameters).

For model recovery (Extended Data Fig. 2d-e), we examined the proportion of times the difference in BIC between the COIN and dual-rate fits favoured the true model class that generated the data.

Simulating existing data sets

We performed COIN model simulations on a diverse set of extant data in Fig. 4 (similarly Extended Data Figs. 5, 8 and 9) in a purely cross-validated manner, such that we used model parameters fitted to participants in our own experiments to make predictions for experiments conducted in other laboratories using other paradigms.

The paradigms in Fig. 4 and Extended Data Fig. 8 were simulated using the 40 sets of parameters fit to our individual participants’ data from both experiments. One hundred simulations (each conditioned on a different noisy state feedback sequence) were performed for each parameter set. The results shown are based on the average of all of these simulations.

The paradigms in Extended Data Fig. 5a-o and Extended Data Fig. 9 were variations of the standard spontaneous recovery paradigm. Therefore, we simulated these paradigms (as well as the paradigm in Extended Data Fig. 5p-s) using the parameters fit to the average spontaneous and evoked recovery data sets. One hundred simulations (each conditioned on a different noisy state feedback sequence) were performed. The results shown are based on the average of these simulations.

Modelling working memory

A working memory task performed after the last P$^-$ trial of a spontaneous recovery paradigm has been shown to interfere with spontaneous recovery, producing an effect that is reminiscent of evoked recovery on the first P^c trial (Extended Data Fig. 9a, Ref. 22). We modelled the effect of the working memory task as selectively abolishing the (working) memory of the responsibilities on the last P$^-$ trial (Extended Data Fig. 9b-d). This means that on the first P^c trial, the predicted probabilities are based on the expected context frequencies (the stationary probabilities).

Modelling visuomotor learning and its explicit and implicit components

In visuomotor rotation experiments, the cursor moves in a different direction to the hand (which is occluded from vision). Hence, visuomotor rotations introduce a discrepancy between the location of the hand as sensed by vision and proprioception. To model this discrepancy, we include a context-specific bias parameter $b^{(c_t)}$ in the state feedback (Eq. 5):

$$y_t = x_t^{(c_t)} + b^{(c_t)} + v_t \qquad v_t \sim \mathcal{N}(0, \sigma_{\mathrm{r}}^2). \tag{19}$$

To support Bayesian inference, we place a normal distribution prior over this parameter:

$$b^{(j)} \mid \mu_{\mathrm{b}}, \sigma_{\mathrm{b}} \sim \mathcal{N}(\mu_{\mathrm{b}}, \sigma_{\mathrm{b}}^2). \tag{20}$$

We set μ_{b} to zero based on the assumption that positive and negative biases are equally probable and σ_{b} to 70^{-1} by hand to match the empirical data in Extended Data Fig. 9e. We extend and modify the inference algorithm accordingly (see Suppl. Inf.).

On each trial, the state feedback was assigned a value of 0 (no rotation trials), $+1$ (P^+ rotation trials) or -1 (P^- rotation trials) plus i.i.d. zero-mean Gaussian observation noise with variance σ_r^2. Visual error-clamp trials (P^c) were modelled in the same way as channel trials (i.e. with state feedback unobserved). Adaptation was modelled as the mean of the predicted *state feedback* distribution (Extended Data Fig. 5q and Extended Data Fig. 9f, dashed pink) plus Gaussian motor noise.

We also modelled an experiment in which an explicit judgement of the perturbation is obtained on every trial, and the implicit component is taken as the difference between adaption and the explicit judgement[23]. We hypothesised that participants have explicit access to the state representing their belief about the visuomotor rotation but do not have access to the bias in the state feedback, which is therefore implicit. Hence, we mapped the state of the context with the highest responsibility on the previous trial (Extended Data Fig. 9h, black line) onto the explicit component and the average bias across contexts weighted by the predicted probabilities (Extended Data Fig. 9j, cyan line) onto the implicit component. Adaptation is then, by definition, the sum of these two components (Extended Data Fig. 9e, solid pink) plus Gaussian motor noise. See Suppl. Inf. for full details.

Data availability

All experimental data are publically available at the Dryad repository (https://doi.org/10.5061/dryad.m63xsj42r). The data include the raw kinematics and force profiles of individual participants on all trials as well as the adaptation measures used to generate the experimental data shown in Fig. 2c,e and Fig. 3d.

Code availability

Code for the COIN model is available at GitHub (https://github.com/jamesheald/COIN).

References

31. Howard, I. S., Ingram, J. N. & Wolpert, D. M. A modular planar robotic manipulandum with end-point torque control. *J. Neurosci. Methods* **181**, 199–211 (2009).
32. Milner, T. E. & Franklin, D. W. Impedance control and internal model use during the initial stage of adaptation to novel dynamics in humans. *J. Physiol.* **567**, 651–664 (2005).
33. Scheidt, R. A., Reinkensmeyer, D. J., Conditt, M. A., Rymer, W. Z. & Mussa-Ivaldi, F. A. Persistence of motor adaptation during constrained, multi-joint, arm movements. *J. Neurophysiol.* **84**, 853–862 (2000).
34. Teh, Y. W., Jordan, M. I., Beal, M. J. & Blei, D. M. Hierarchical Dirichlet processes. *J. Amer. Stat. Assoc.* **101**, 1566–1581 (2006).
35. Fox, E. B., Sudderth, E. B., Jordan, M. I. & Willsky, A. S. An HDP-HMM for systems with state persistence. In *Proc. 25th Int. Conf. Machine Learning*, 312–319 (2008).
36. Teh, Y. W. Dirichlet processes. In *Encyclopedia of Machine Learning* (Springer, 2010).
37. Carvalho, C. M., Johannes, M. S., Lopes, H. F. & Polson, N. G. Particle learning and smoothing. *Stat. Sci.* **25**, 88–106 (2010).
38. Bernardo, J. *et al.* Particle learning for sequential Bayesian computation. *Bayesian Statistics 9* **9**, 317 (2011).

39. Houlsby, N. *et al.* Cognitive tomography reveals complex, task-independent mental representations. *Curr. Biol.* **23**, 2169–2175 (2013).
40. Acerbi, L. & Ji, W. Practical Bayesian optimization for model fitting with bayesian adaptive direct search. In *Adv. Neural Inf. Proc. Sys.*, 1836–1846 (2017).
41. Jeffreys, H. *The theory of probability* (OUP Oxford, 1998).
42. Li, J., Wang, Z. J., Palmer, S. J. & McKeown, M. J. Dynamic Bayesian network modeling of fMRI: a comparison of group-analysis methods. *Neuroimage* **41**, 398–407 (2008).

Acknowledgements

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 726090 to M.L.), the Wellcome Trust (Investigator Awards 212262/Z/18/Z to M.L. and 097803/Z/11/Z to D.M.W.), Royal Society (Noreen Murray Professorship in Neurobiology to D.M.W), National Institutes of Health (R01NS117699 and U19NS104649 to D.M.W.) and Engineering and Physical Sciences Research Council (studentship to J.B.H). For the purpose of open access, the authors have applied a CC-BY public copyright licence to any author accepted manuscript version arising from this submission. We thank J. Ingram for technical support and G. Hennequin for discussions.

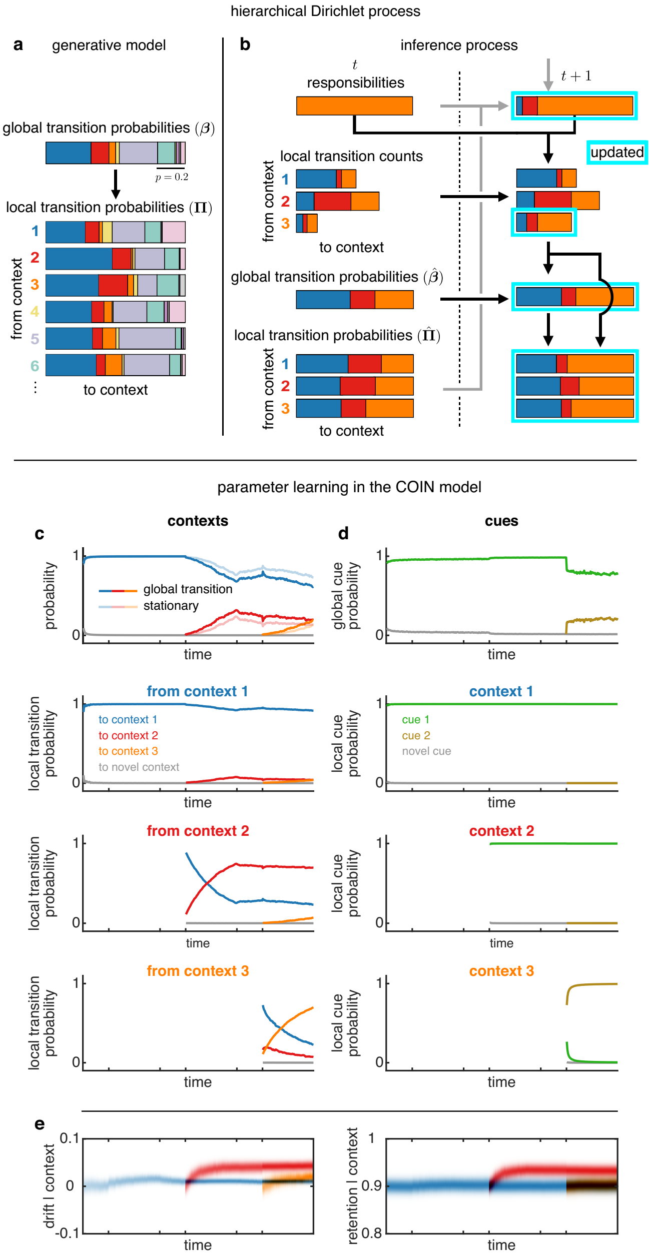
Authors' contributions

J.B.H. developed the model, implemented the model, performed the experiments, analysed the data and performed simulations. J.B.H. and D.M.W. designed the behavioural experiments. All authors were involved in the conceptualisation of the study, developed techniques for analysing the model, interpreted results and wrote the paper.

Competing interests

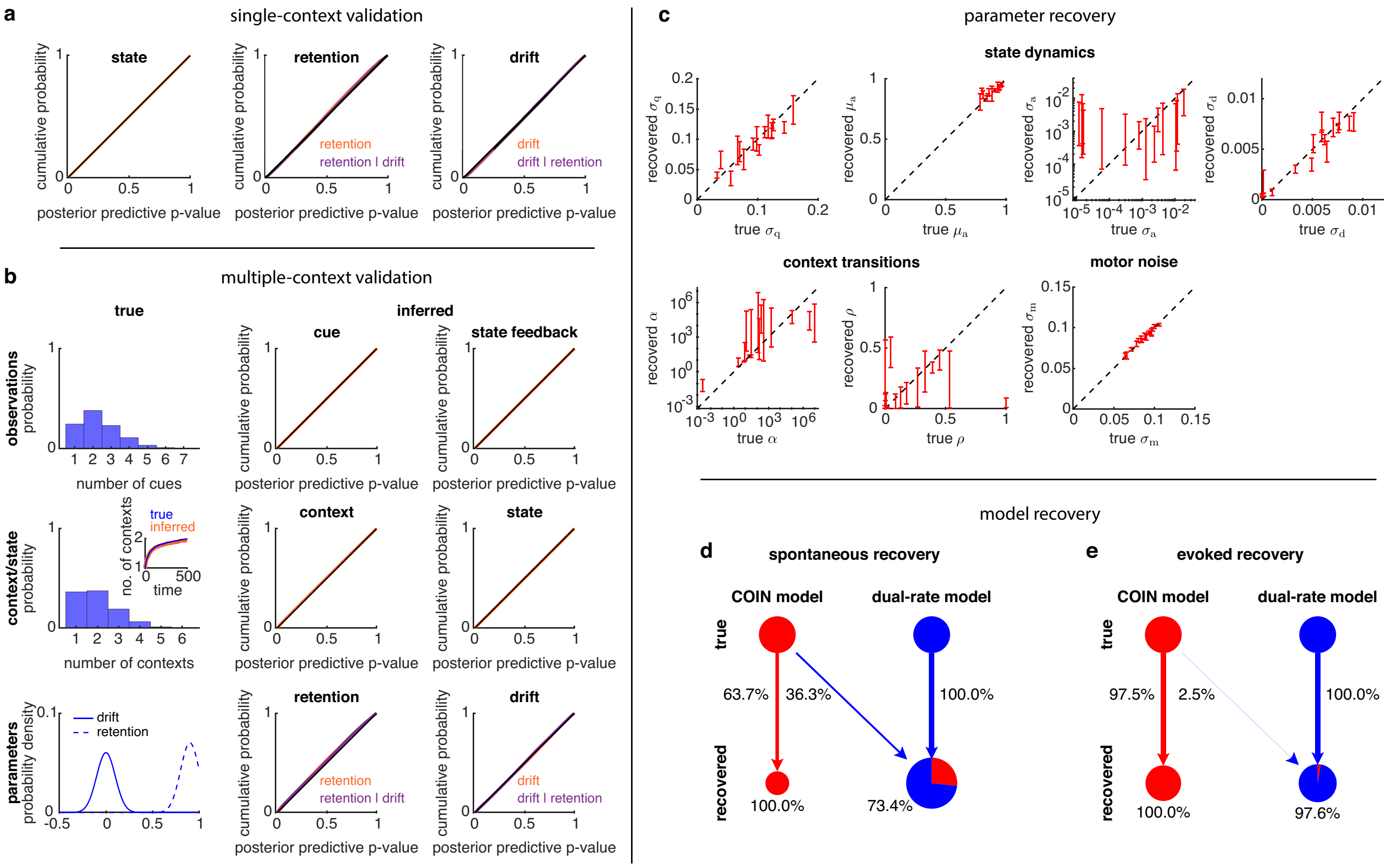
The authors have no competing interests.

Extended data figures



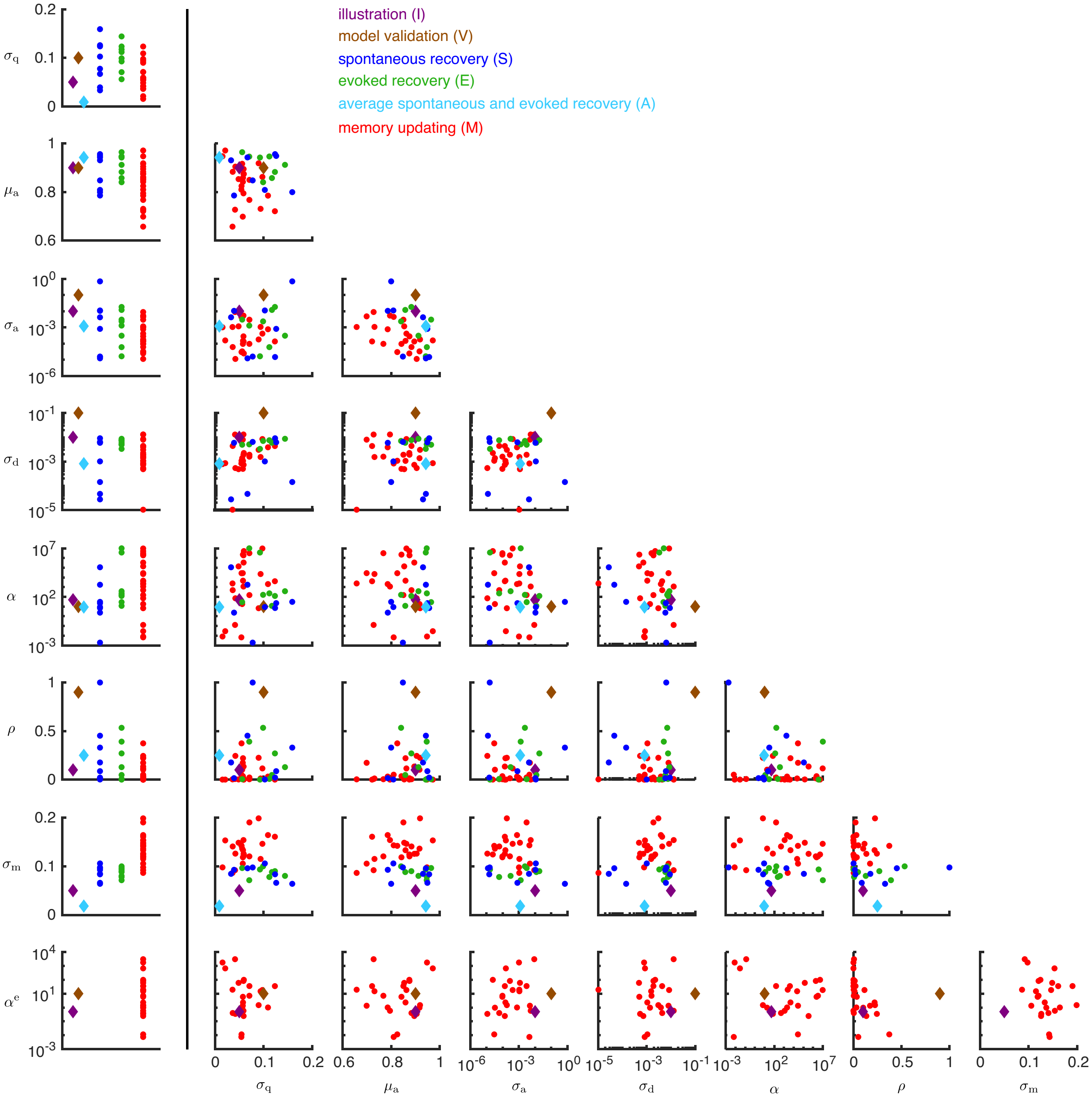
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Additional details of the COIN model (related to Fig. 1). a-b, Hierarchy and generalisation in contextual inference. a, Local transition probabilities are generated in two steps via a hierarchical Dirichlet process. In the first step (top), an infinite set of global transition probabilities $\boldsymbol{\beta}$ are generated via a stochastic stick-breaking process (see Suppl. Inf.). Probabilities are represented by the width of bar segments with different colours indicating different contexts. In the second step (bottom), for each context ('from context'), local transition probabilities to each other context ('to context') are generated (a row of $\boldsymbol{\Pi}$) via a stochastic Dirichlet process and are equal to the global probabilities in expectation (bar a self-transition bias, which we set to zero here for clarity). (An analogous hierarchical Dirichlet process, not shown, is used to generate the global and local cue probabilities.) **b,** Contextual inference updates both the global and local transition probabilities. Context transition counts are maintained for all from-to pairs of known contexts and get updated based on the contexts inferred on two consecutive time points (responsibilities at time points t and $t+1$). These updated context transition counts are used to update the inferred global transition probabilities $\hat{\boldsymbol{\beta}}$. The updated global transition probabilities and context transition counts produce new inferences about the inferred local transition probabilities $\hat{\boldsymbol{\Pi}}$. Note that although the model infers full (Dirichlet) posterior distributions over both the global and local transition probabilities, for clarity here we only show the means of these posterior distributions (indicated by the hat notation). In the example shown, only row 3 of the context transition counts is updated (as context 3 has an overwhelming responsibility at time t), but all rows of the local transition probabilities are updated due to the updating of the global transition probabilities (if the model were non-hierarchical, there would be no global transition probabilities, and so the local transition probabilities would only be updated for context 3 via the updated context transition contexts). Thus inferences about transition probabilities generalise from one context (here context 3) to all other contexts (here contexts 1 and 2) due to the hierarchical nature of the generative model. Note that when a novel context is encountered for the first time, its local transition probabilities are initialised based on $\hat{\boldsymbol{\beta}}$, thus allowing well-informed inferences about transitions to be drawn immediately. **c-e, Parameter inference in the COIN model for the simulation shown in Fig. 1c-f.** In addition to inferring states and contexts, the COIN model also infers transition (**c**) and cue (**d**) probabilities, as well as the parameters of context-specific state dynamics (**e**). **c,** Transition probabilities. Top: Estimated global transition probabilities (solid lines) to each known context (line colours) and the novel context (grey). Pale lines show estimated stationary probabilities of the same contexts representing the expected proportion of time spent in each context given the current estimate of the local transition probabilities (below). Bottom three panels: estimated local transition probabilities from each context (colours as in top panel). **d,** Estimated global (top panel) and local cue probabilities for the three known contexts (bottom three panels) and cues (line colours). Although the model infers full (Dirichlet) posterior distributions over both transition (**c**) and cue probabilities (**d**), for clarity here we only show the means of these posterior distributions. **e,** Posterior distribution of drift (left) and retention parameters (right) for the three known contexts (colours as in **c**, novel context not shown for clarity). Although the model infers the joint distribution of the drift and retention parameters for each context, for clarity here we show the marginal distribution of each parameter separately. Note that drift and retention are estimated to be larger for the red context that is associated with the largest perturbation.

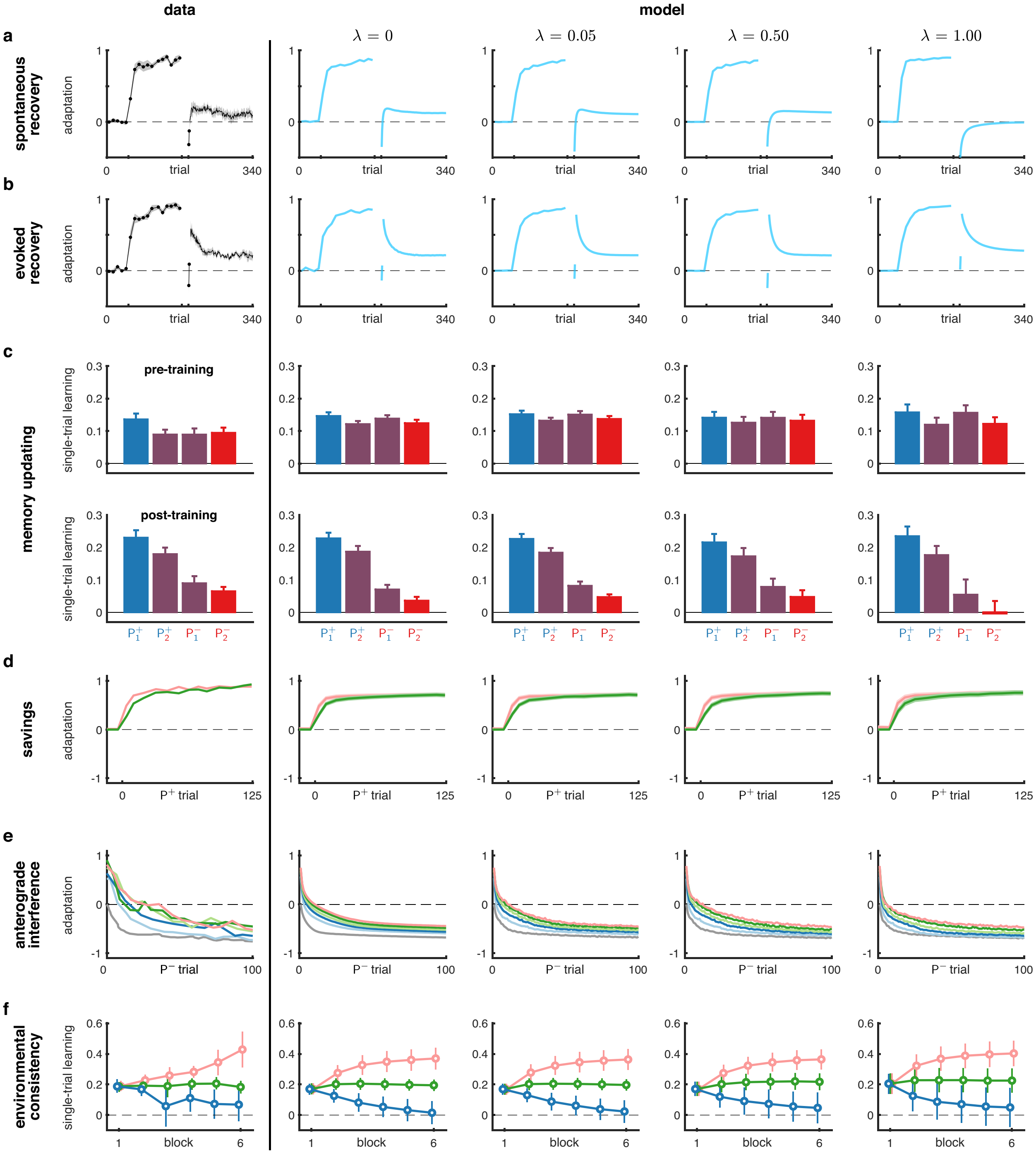


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Validation of the COIN model. a, Validation of the inference algorithm of the COIN model with a single context. We computed inferences in the COIN model with a single context based on synthetic observations (state feedback) generated by its generative model (Fig. 1a). Plots show the cumulative distributions of posterior predictive p-values of the state variable (left), and the parameters governing its dynamics (retention, middle; drift, right). The posterior predictive p-value is computed by evaluating the cumulative distribution function of the model's posterior over the given quantity at the true value of that quantity (as defined by the generative model). Empirical distributions of posterior predictive p-value were collected across 4000 simulations (with different true state dynamics parameters), with 500 time steps in each simulation (during which the true state changes, but the state dynamics parameters are constant). Note that although true state dynamics parameters do not change during a simulation, inferences in the model about them will still generally evolve, and so a new posterior p-value is generated in each time step even for these quantities. If the model implements well-calibrated probabilistic inference under the correct generative model, all these empirical distributions should be uniform. This is confirmed by all cumulative distributions (orange and purple curves) approximating the identity line (black diagonal). Orange curves show posterior predictive p-values under the corresponding marginals of the model's posterior. To give additional information about the model's joint posterior over state dynamics parameters, we also show the posterior predictive p-value (cumulative) distribution of each parameter conditioned on the true value of the other one (purple curves). **b, Validation of the inference algorithm of the COIN model with multiple contexts.** Simulations as in **a** but with additional synthetic observations (sensory cues) and multiple contexts allowed both during data generation and inference. Empirical distributions of posterior predictive p-value were collected across 2000 simulations (with different true retention and drift parameters), with 500 time steps in each simulation (during which not only states evolve but also contexts transition, and sometimes novel contexts are created). Left column shows the true distributions of sensory cues, contexts and parameters. Inset shows the growth of the number of contexts over time both during generation (blue) and inference (orange). Middle and right columns show the cumulative probabilities of the posterior predictive p-values (pooled across data sets and time steps) for the observations (top row), contexts and state (middle row) and parameters (bottom row). To calculate the posterior predictive p-values for the context, inferred contexts were relabelled by minimising the Hamming distance between the relabelled context sequence and the true context sequence (see Suppl. Inf.). For the parameters, the posterior predictive p-values were calculated with respect to both the marginal distributions (retention and drift) and the conditional distributions (retention $|$ drift, and drift $|$ retention) as in **a**. The cumulative probability curves approximate the identity line (thin black line) showing that the inferred posterior probability distributions are well calibrated. **c, Parameter recovery in the COIN model related to Fig. 2.** Plots show the COIN model parameters that were recovered (y-axes) from fits to 10 synthetic data sets generated with the COIN model parameters (true, x-axes) obtained from the fits to each participant in the spontaneous (n = 8) and evoked (n = 8) recovery experiments (Extended Data Fig. 3). Vertical bars show the interquartile range of the recovered parameters for each participant. While several parameters are recovered with good accuracy ($\sigma_\mathrm{q}, \mu_\mathrm{a}, \sigma_\mathrm{d}, \sigma_\mathrm{m}$), others are not ($\alpha$, and in particular σ_a and ρ). We expect that with richer paradigms and larger data sets, all parameters would be recovered accurately. Most importantly, despite partial success with recovering individual parameters, model recovery shows that recovered parameter sets taken as a whole can be used to accurately identify whether data was generated by the dual-rate or COIN model (**d**). Note that we make no claims about individual parameters in this study as our focus is on model class recovery. **d-e, Model recovery for spontaneous (d) and evoked recovery experiments (e) related to Fig. 2.** Synthetic data sets were generated using one of two models (COIN model, red; dual-rate model, blue). Parameters used for each model were those obtained from the fits to each participant in the spontaneous (n = 8) and evoked (n = 8) recovery experiments (Extended Data Fig. 3) – i.e. for the COIN model, these were the same synthetic data sets as those used in **c**. Then, the same model comparison method that we used on real data (Fig. 2c, e, insets) was used to recover the model that generated each synthetic data set (see Methods). Arrows connect true models (used to generate synthetic data, disks on top) to models that were recovered from their synthetic data (pie-chart disks at bottom). Arrow colour indicates identity of recovered model, arrow thickness and percentages indicate probability of recovered model given true model. Bottom disk sizes and pie-chart proportions show total probability of recovered model and posterior probability of true model given recovered model (assuming a uniform prior over true models), respectively, with percentages specifically indicating posterior probability of the correct model. These results show that the model recovery process is generally very accurate and actually biased against the COIN model in favour of the dual-rate model.

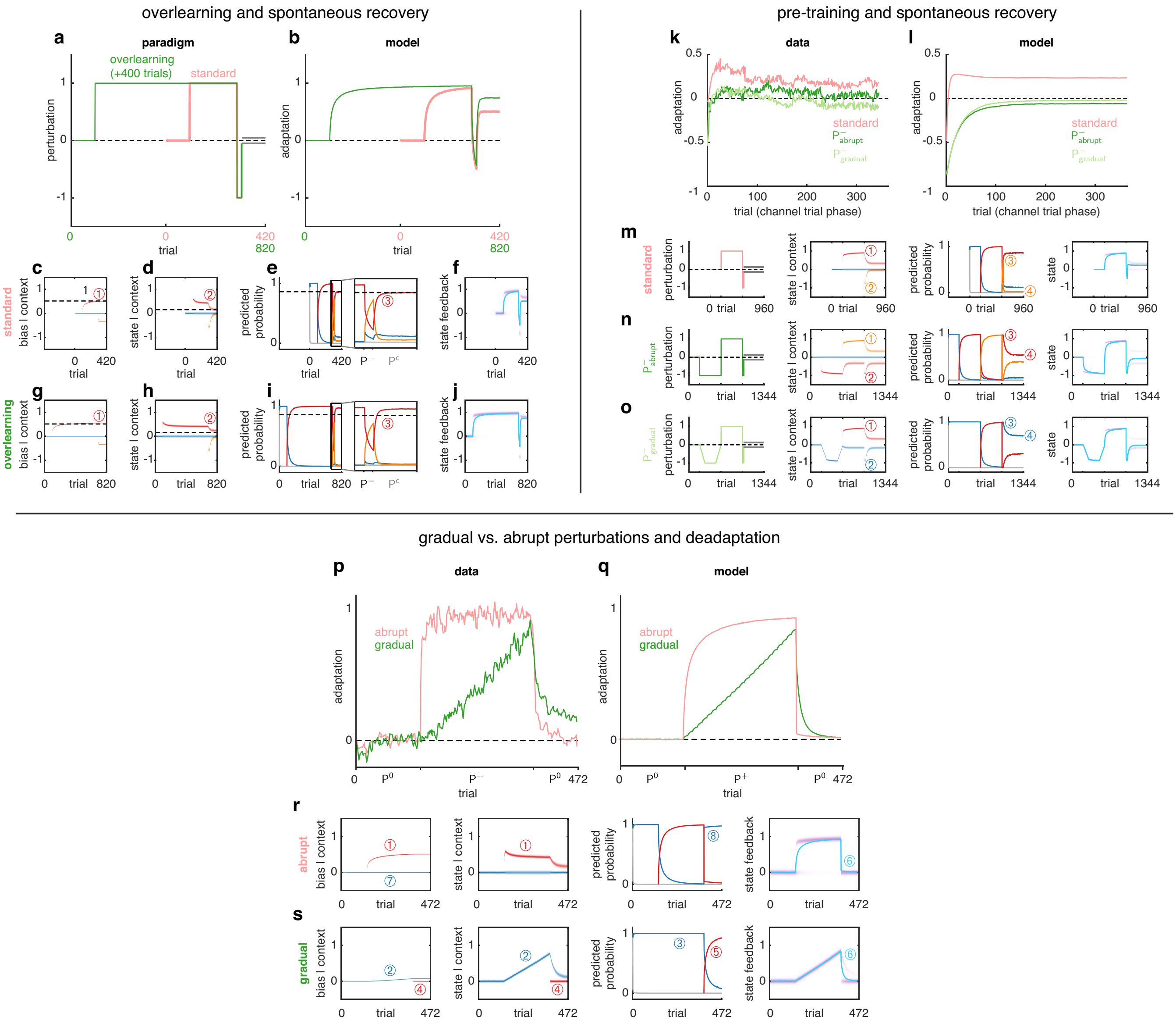


Extended Data Fig. 3 | COIN model parameters. Left column: Parameters for illustrating the COIN model (**I**: purple), model validation (**V**: brown) and fits to individuals in the spontaneous (**S**: blue) and evoked (**E**: green) recovery experiments, to the average of both groups (**A**: cyan), and individuals in the memory-updating experiment (**M**: red). Right: scatter plots for all pairs of parameters for the six groups. The overlap of data points suggest parameters are similar across experiments. σ_q: process noise s.d. (Eq. 3); μ_a, σ_a: prior mean and s.d. for context-specific state retention factors (Eq. 10); σ_d: prior s.d. for context-specific state drifts (Eq. 10); α: concentration of local transition probabilities (Eq. 8); ρ: self-transition bias parameter (Eq. 18); σ_m: motor noise s.d. (Eq. 17); α^e: concentration of local cue probabilities (Eq. 9). Parameters used in the figures is as follows. **I**: Fig. 1 and Extended Data Fig. 1c-e. **V**: Extended Data Fig. 2a-b. **S**: Fig. 2c, Extended Data Fig. 6f (column 1) and Extended Data Fig. 2d. **E**: Fig. 2e, Extended Data Fig. 6f (column 3) and Extended Data Fig. 2e. **S** & **E**: Extended Data Fig. 2c. **A**: Fig. 2b & d, Extended Data Fig. 5 and Extended Data Fig. 9 (bias added for visuomotor rotation experiments: Extended Data Fig. 5a-j,p-s and Extended Data Fig. 9e-l). **M**: Fig. 3 and Extended Data Fig. 7a-d. **S**, **E** & **M**: (all parameters, but α^e): Fig. 4 and Extended Data Fig. 8. The robustness analyses (Extended Data Fig. 4) used perturbed versions of the same parameters as the corresponding unperturbed simulations. To reduce the number of free parameters in the model, we set the parameters of the hierarchical Dirichlet process that determine the expected effective number of contexts or cues, γ (Eq. 7) and γ^e (Eq. 9), respectively, both to 0.1, the prior mean for context-specific state drifts, μ_d, to zero (Eq. 10), and the standard deviation of the sensory noise, σ_s, to 0.03 when fitting or simulating the model, with the variance of the observation noise (Eqs. 5 and 19) being set to $\sigma_\mathrm{r}^2 = \sigma_\mathrm{s}^2 + \sigma_\mathrm{m}^2$. For visuomotor rotation experiments (Extended Data Fig. 5a-j,p-s and Extended Data Fig. 9e-l), we set the mean of the prior of the bias μ_b to zero (Eq. 20), and its s.d. σ_b to 70^{-1}.



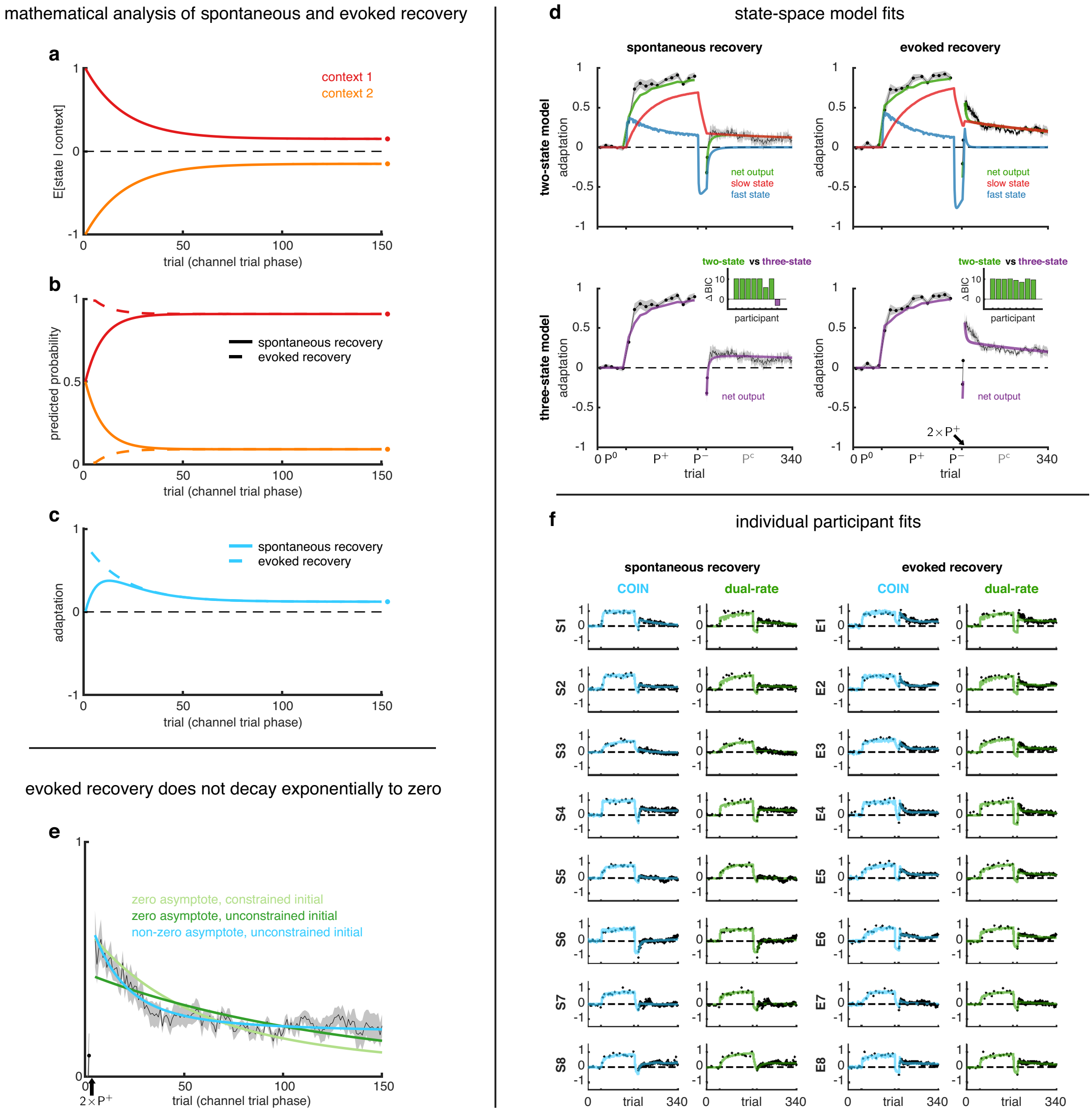
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Robustness analysis of the main COIN model results. To test how robust the behaviour of the COIN model is, we added noise to the parameters fit to the individual participants in the spontaneous and evoked recovery, and memory updating experiments and re-simulated the paradigms in Figs. 2 to 4: spontaneous recovery (**a**), evoked recovery (**b**), memory updating (**c**), savings (**d**), anterograde interference (**e**), and environmental consistency (**f**). For each experiment, we simulated the COIN model for the same participants as in Figs. 2 to 4 but perturbed each participant's parameter values. That is, for each parameter (suitably transformed to be unbounded) we calculated the standard deviation across participants (relevant for the given paradigm or set of paradigms) and then perturbed each participant's (transformed) parameter by zero-mean Gaussian noise whose standard deviation was a fraction (λ = 0, 0.05, 0.5, or 1.0) of this empirical standard deviation, after which we used the inverse transform to obtain the actual parameter used in these perturbed simulations. For parameters that are constrained to be non-negative (σ_q, σ_a, σ_d, α, α^e, σ_m), we used a logarithmic transformation, whereas for parameters constrained to be on the unit interval (μ_a, ρ), we used a logit transformation. Column 1: experimental data (plotted as in Figs. 2 to 4). Columns 2-5: output of the COIN model for different amounts of noise added to the parameters. Note that the simulations were not conditioned on the actual adaptation data of individual participants (in contrast to the original simulations of Figs. 2 and 3) because these data are not available for the experiments shown in Fig. 4 (for which the original simulations were already performed using this 'open-loop' simulation approach). The robustness analysis shows that most predictions of the COIN model are robust to changes in the parameters, and only start to deviate for large parameter changes (λ = 1) in some of their quantitative details (such as the magnitude of spontaneous recovery). Note that $\lambda = 1$ leads to changes in parameters that are of the same magnitude as randomly shuffling the parameters across participants.



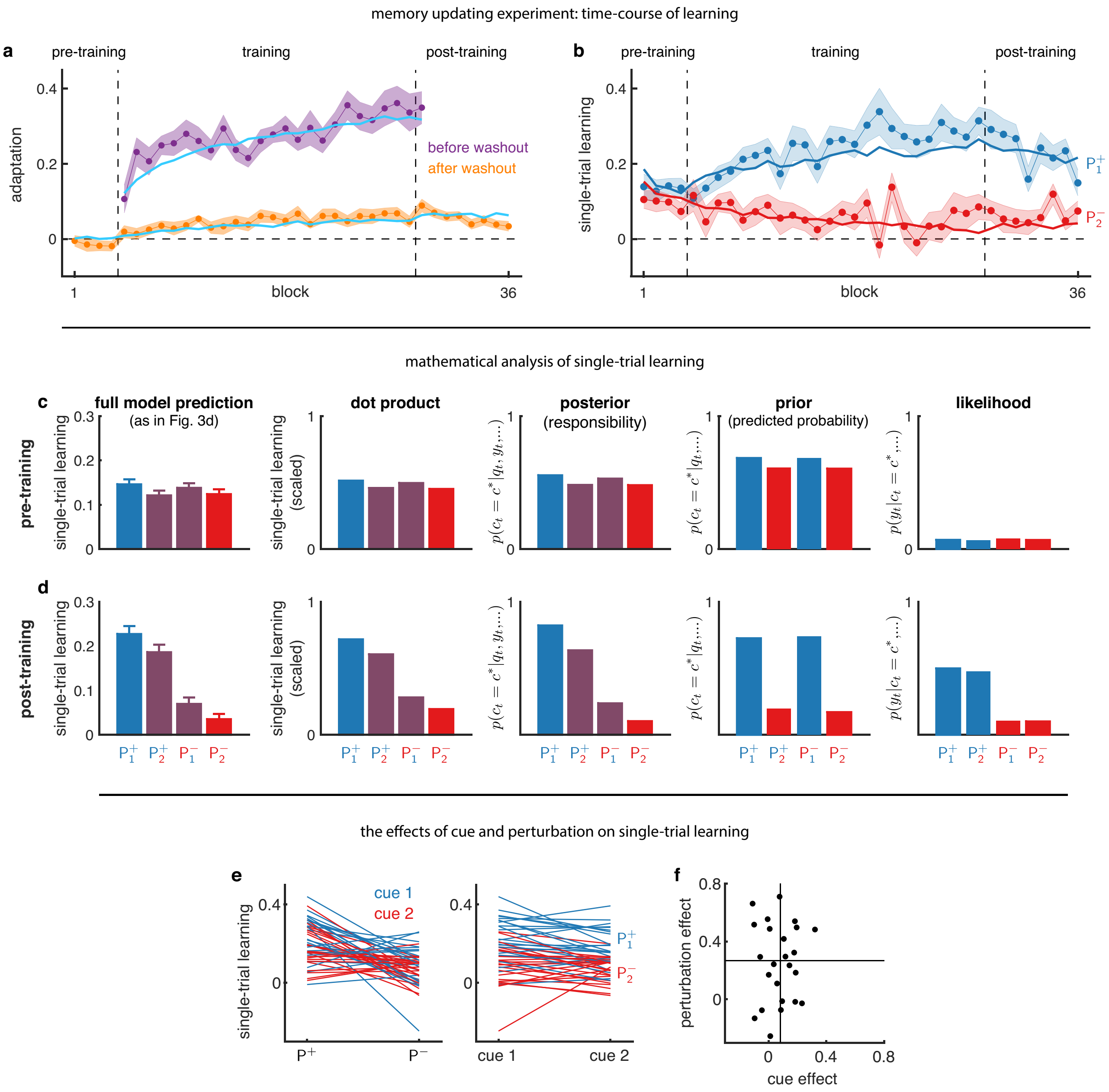
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | History dependence of contextual inference. a-**j**, **Contextual inference underlies the elevated level of spontaneous recovery after 'overlearning'. a,** Spontaneous recovery paradigm for visuomotor learning in which the length of the exposure (P^{+}) phase is tripled from 200 trials ('standard' paradigm, pink) to 600 trials ('overlearning' paradigm, green). For comparison, paradigms are aligned to the end of the exposure phase. **b,** Adaptation in the COIN model for the standard and overlearning paradigms (same parameters as in Fig. 2b & d but with the addition of a bias parameter; see Suppl. Inf. and also Extended Data Fig. 3, parameter set **A**). Adaptation corresponds to reach angle normalised by the size of the experimentally-imposed visuomotor rotation. Note elevated level of spontaneous recovery after overlearning compared to the standard paradigm, qualitatively matching visuomotor learning data in Fig. 4A of Ref. 13. **c**-**f**, Internal representations of the COIN model for the standard paradigm. Inferred bias (**c**) and predicted state (**d**) distributions for each context (colours). **e,** Predicted probabilities of each context (with zoomed view starting from near the end of P^{+} exposure), colours as in **c**-**d**, grey is novel context as in Fig. 1f. **f**, Predicted state feedback (predicted state plus bias) distribution (purple), which is a mixture of the individual contexts' predicted state feedback distributions (not shown) weighted by their predicted probabilities (**e**). Total adaptation (cyan line) is the mean of the predicted state feedback distribution. **g**-**j**, same as **c**-**f** for the overlearning paradigm. For comparison, the dashed horizontal lines in both paradigms show the final level of each variable for the red context in the standard paradigm. Note that overlearning leaves inferences about biases and states largely unchanged (compare 1 in **c** & **g** and 2 in **d** & **h**) but leads to higher predicted probabilities of the P^{+} context (red) in the channel-trial phase (compare 3 in **e** & **i**) reflecting the true statistics of the experiment in which P^{+} occurred more frequently. In turn, this makes the P^{+} bias and state contribute more to total adaptation in the channel-trial phase, thus explaining higher levels of spontaneous recovery. Therefore, differences between conditions are explained by contextual inference rather than by differences in bias or state inferences. The results are qualitatively similar when simulated as a force-field paradigm (i.e. without bias, not shown). **k**-**o**, **Contextual inference underlies reduced spontaneous recovery following pre-training with P^{-}. k**, Adaptation in the channel-trial phase of a typical spontaneous recovery paradigm (standard, pink, as in Fig. 2b) and two modified versions of the paradigm in which the P^{+} phase is preceded by a P^{-} (pre-training) phase in which P^{-} is either introduced and removed abruptly ($\mathrm{P}^{-}_{\mathrm{abrupt}}$, dark green) or gradually ($\mathrm{P}^{-}_{\mathrm{gradual}}$, light green). Data reproduced from Ref. 14. **l-o,** Simulation of the COIN model for the same paradigms (same parameters as in Fig. 2b and d; Extended Data Fig. 3, parameter set **A**), plotted as in Fig. 2b-c. In each paradigm, contexts are coloured according to their order of instantiation during inference (blue→red→orange). Note that pre-training with P^{-} (either abrupt or gradual) leaves inferences about states within each context largely unchanged at the beginning of the channel-trial phase (compare corresponding numbers 1-2 in column 2 across **m**-**o**). However, the pre-training leads to higher predicted probabilities of the P^{-} context initially (compare number 3 in **m** to 3 in **n** & **o**) and throughout the channel-trial phase (compare number 4 across **m**-**o**) reflecting the true statistics of the experiment in which P^{-} occurred more frequently (compare column 1 across **m**-**o**). In turn, this makes the P^{-} state contribute more to total adaptation, thus explaining the reduction in both the initial and final levels of adaptation during the channel-trial phase in the $\mathrm{P}^{-}_{\mathrm{abrupt}}$ and $\mathrm{P}^{-}_{\mathrm{gradual}}$ groups. Therefore, as in Fig. 4, differences between conditions are explained by contextual inference rather than state inference. **p**-**s, Contextual inference underlies slower deadaptation following a gradually-introduced perturbation. p,** Adaptation (normalised reach angle, as in **b**) in a paradigm in which a visuomotor rotation is introduced abruptly (pink) or gradually (green) and then removed abruptly. Data reproduced from Ref. 17. **q**-**s**, Simulation of the COIN model on the abrupt (**q**, pink, and **r**) and gradual (**q**, green, and **s**) paradigms (same parameters as in Fig. 2b and d but with the addition of a bias parameter; Extended Data Fig. 3, parameter set **A**) plotted as in **b**-**j**. Note that contexts are coloured according to their order of appearance during inference (blue→red). In response to the abrupt introduction of the P^{+} perturbation, a new memory is created (1). In contrast, the gradual introduction of the P^{+} perturbation prevents the creation of a new memory, thus requiring changes in the inferred bias and state of the original memory associated with P^{0} (2, blue context) to account for the slowly increasing perturbation. Therefore, the 'blue' context is inferred to be active throughout the exposure phase (3) and becomes associated with a P^{+}-like state. However, at the beginning of the abruptly introduced post-exposure (P^{0}) phase, a new memory is created (4) which has a low initial probability that can only be increased by repeated experience with P^{0} (5). This leads to slower deadaptation in the post-exposure phase compared to the abrupt paradigm (6), in which the original context associated with P^{0} (blue) is protected (7) and can be reinstated quickly (8) as the P^{0} self-transition probability has been learned to be higher during the pre-exposure phase. Note that the smaller errors caused by the gradual perturbation relative to the abrupt condition are better accounted for by an error in the state rather than an error in the bias, and therefore the state is updated more than the bias. The results are qualitatively similar when simulated as a force field paradigm (without bias, not shown).



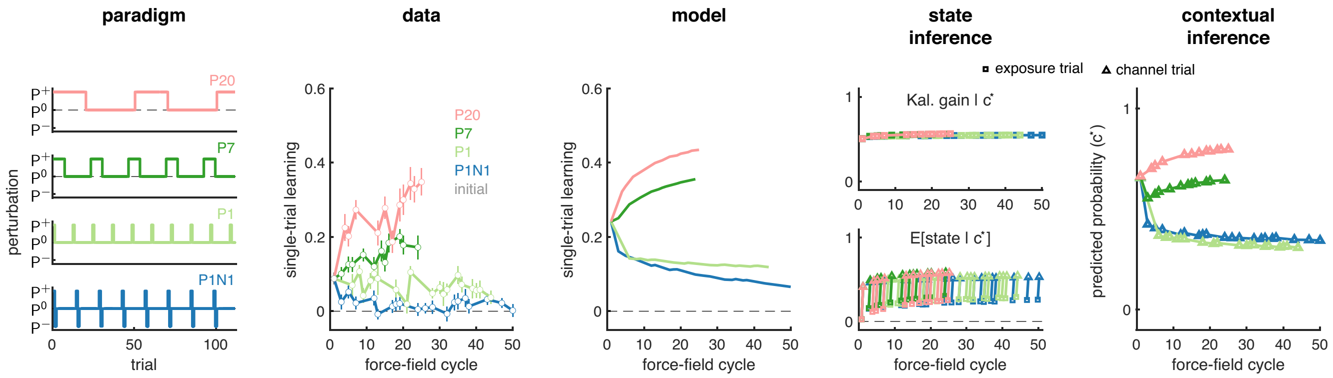
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Additional analyses of spontaneous and evoked recovery related to Fig. 2. a-c, Mathematical analysis of spontaneous and evoked recovery. The channel-trial phase of spontaneous and evoked (after the two P^{+} trials) recovery simulated in a simplified setting (Suppl. Inf.) with two contexts that are initialised to have equal but opposite state estimates (**a**) and equal (spontaneous recovery, solid) or highly unequal (evoked recovery, dashed) predicted probabilities (**b**). For the two contexts, the retention parameters are assumed to be constant and equal, and the drift parameters are assumed to be constant, of the same magnitude but opposite sign. Mean adaptation (**c**), which in the COIN model is the average of the state estimates (**a**) weighted by the corresponding context probabilities (**b**), shows the classic pattern of spontaneous recovery (solid, cf. Fig. 2b-c) and the characteristic abrupt rise of evoked recovery (dashed, cf. Fig. 2d-e). Note that although in the full model, state estimates are different between evoked and spontaneous recovery following the two P^{+} trials, here we assumed they are the same (no separate solid and dashed lines in **a**) for simplicity and to demonstrate that the difference in mean adaptation between the two paradigms (**c**) can be accounted for by differences in contextual inference alone (**b**, cf. Fig. 2b and d, top right insets). Circles on the right show steady-state values of inferences and the adaptation. Note that in both paradigms, adaptation is predicted to decay to a non-zero asymptote (see also **e**). **d, State-space model fits to adaptation data from the spontaneous and evoked recovery groups.** Solid lines show the mean fits across participants of the two-state model (5 parameters, top row) and the three-state model (7 parameters, bottom row) to the spontaneous recovery (left column) and evoked recovery (right column) data sets. Mean ± SEM adaptation on channel trials shown in black (same as in Fig. 2c and e). Insets show differences in BIC (nats) between the two-state model and the three-state model for individual participants (positive values in green indicate evidence in favour of the two-state model, and negative values in purple indicate evidence in favour of the three-state model). At the group level, the two-state model was far superior to the three-state model (Δ group-level BIC of 64.2 and 78.4 nats favour of the two-state model for the spontaneous and evoked recovery groups, respectively). Individual states are shown for the two-state model (top, blue and red). Both the fast and slow processes adapt to P^{+} during the extended initial learning period. The P^{-} phase reverses the state of the fast process, but not of the slow process, so that they cancel when summed resulting in baseline performance. Spontaneous recovery during the P^{c} phase is then explained by the fast process rapidly decaying, revealing the state of the slow process that has remained partially adapted to P^{+}. Note that this explanation is because in multi-rate models all processes contribute equally to the motor output at all times. This is fundamentally different from the expression and updating of multiple context-specific memories in the COIN model, which are dynamically modulated over time according to ongoing contextual inference. **e, Evoked recovery does not decay exponentially to zero.** According to the COIN model, adaptation in the channel-trial phase of evoked recovery can be approximated by exponential decay to a non-zero (i.e. positive) asymptote (**a-c**, Fig. 2e, Suppl. Inf.). To test this prediction, we fit an exponential function that either decays to zero (light and dark green) or decays to a non-zero (constrained to be positive) asymptote (cyan) to the adaptation data of individual participants in the evoked recovery group after the two P^{+} trials (black arrow). The two zero-asymptote models differ in terms of whether they are constrained to pass through the datum on the first (channel) trial (light green) or not (dark green). The mean fits across participants for the models that decay to zero (green) fail to track the mean adaptation (black, ± SEM across participants), which shows an initial period of decay followed by a period of little or no decay. The mean fit for the model that decays to a non-zero asymptote (cyan) tracks the mean adaptation well and was strongly favoured in model comparison (Δ group-level BIC of 944.3 and 437.7 nats compared to the zero-asymptote fits with constrained and unconstrained initial values, respectively). Note that fitting to individual participants excludes the confound of finding a more complex time course (e.g. one with non-zero asymptote) only due to averaging across participants that each show a different simple time course (e.g. all with zero asymptote but different time constants). **f, COIN and dual-rate model fits for individual participants in the spontaneous and evoked recovery groups.** Data and model predictions are shown for individual participants as in Fig. 2c and e for across-participant averages. Participants in the S and E groups are ordered by decreasing BIC difference between the dual-rate and COIN model (i.e. S1's and E1's data most favour the COIN model), as in insets of Fig. 2c and e. Note that the COIN model can account for much of the heterogeneity of spontaneous (e.g. from large in S1 to minimal in S6) and evoked recovery (e.g. from large in E1 to minimal in E7).

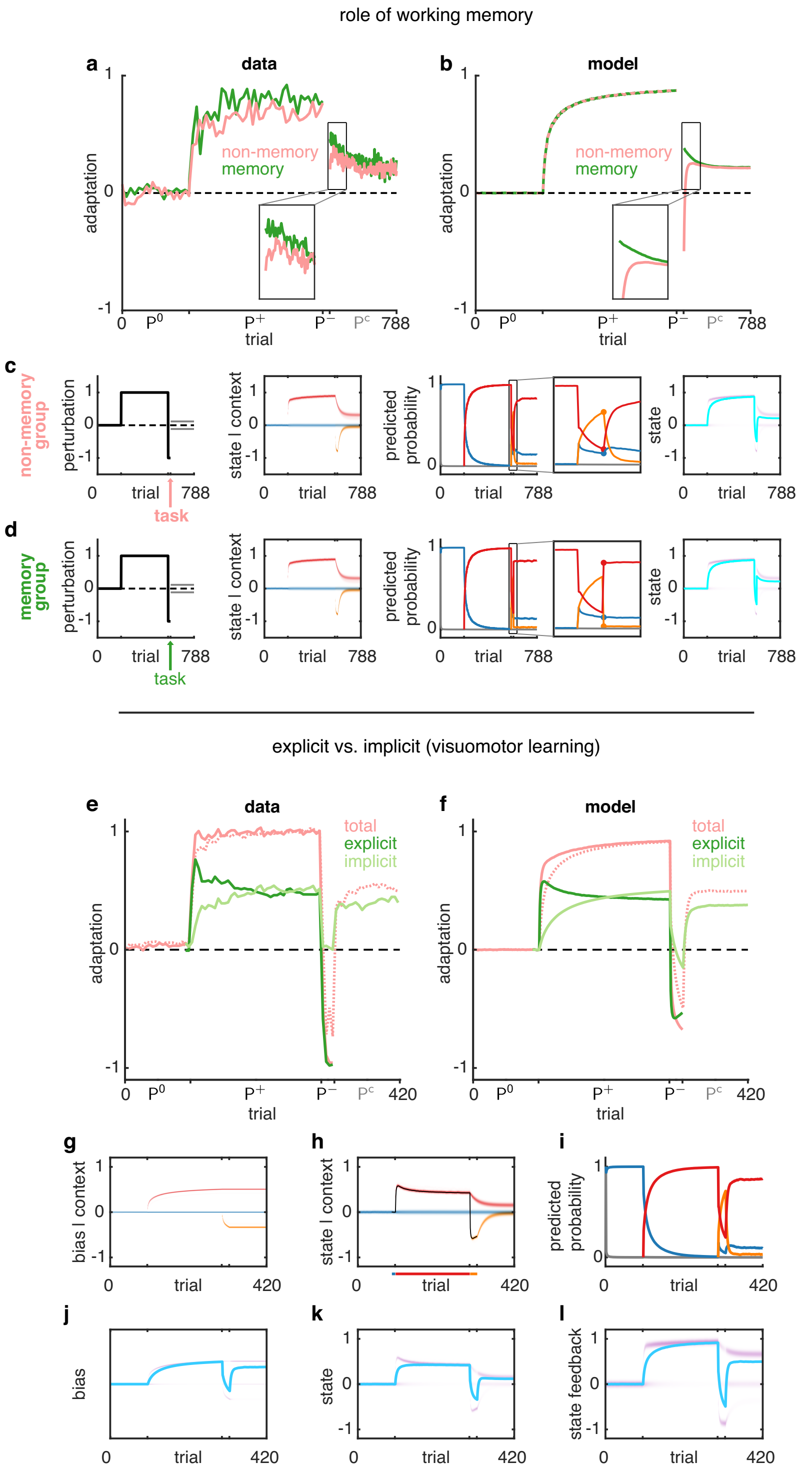


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Additional analyses of memory updating experiment (related to Fig. 3). **a**-**b**, **Memory updating experiment: time-course of learning.** **a**, Adaptation on channel trials at the end of each block of force-field trials in the training phase (purple), which occur before P^0 washout trials, and on the first channel trial of triplets within each block (orange), which occurs after P^0 washout trials. Data is mean ± SEM across participants and lines show mean of COIN model fits (8 parameters, Extended Data Fig. 3). **b**, Single-trial learning on triplets that were consistent with the training contingencies. Data (mean ± SEM across participants) with mean of COIN model fits across participants. Positive learning reflects changes in the direction expected based on the force field of the exposure trial (an increase following P$^+$ and a decrease following P$^-$). **c**-**d**, **Mathematical analysis of single-trial learning.** Single-trial learning in the COIN model (column 1) for the four cue-perturbation triplets in the pre-training phase (**c**) and the post-training phase (**d**) in the memory updating experiment. The COIN model was fit to each participant and model fits are shown as mean ± SEM (single-trial learning) or mean (dot product, posterior, prior and likelihood) across n = 24 participants. Single-trial learning (column 1) is approximately proportional to a dot product (column 2) between the vector of posterior context probabilities (responsibilities) on the exposure trial of the triplet and the vector of predicted context probabilities on the subsequent channel trial (see Suppl. Inf. for derivation). This dot product can be further approximated by collapsing the vector of predicted probabilities to a one-hot vector, i.e. by the responsibility $p(c_t = c^*|q_t, y_t, ...)$ (column 3) of the context that is predominantly expressed on the subsequent channel trial (c^*, the context with the highest predicted probability), where ... denotes all observations before time t (as in Fig. 1). This responsibility is proportional to a product of two terms. The first term is the prior context probability $p(c_t = c^*|q_t, ...)$ (column 4), i.e. the predicted context probability before experiencing the perturbation (as in Fig. 1f$_1$), which is already conditioned on the sensory cue visible from the outset of the trial. The second term expresses the likelihood of the state feedback in that context $p(y_t|c_t = c^*, ...)$ (column 5). As prior to learning neither cues nor feedback are yet consistently associated with a particular context, the COIN model predicts that the prior and likelihood, and thus total single-trial learning should all be largely uniform across contexts before training. **e**-**f**, **The effects of cue and perturbation on single-trial learning in individual participants.** **e**, Single-trial learning (post-training) shown as a function of perturbation separated by cue (left) or as a function of cue separated by perturbation (right) for each participant (lines). Note a significant effect for both the perturbation and the cue. **f**, Scatter plot of cue effect ($\mathrm{P}_1^+ + \mathrm{P}_1^- - \mathrm{P}_2^+ - \mathrm{P}_2^-$) against perturbation effect ($\mathrm{P}_1^+ + \mathrm{P}_2^+ - \mathrm{P}_1^- - \mathrm{P}_2^-$) for each participant (dots). Solid lines show medians of corresponding effects. Note the lack of anti-correlation between two effects.



Extended Data Fig. 8 | Additional analysis of the effect of environmental consistency on single-trial learning related to Fig. 4c. Columns 1 & 2: experimental paradigm and data replotted from Ref. 5. Participants experienced repeating cycles of P$^+$ trials of varying lengths (column 1: 20 P$^+$ trials in P20, 7 in P7, 1 in P1 and 1 followed by 1 P$^-$ trial in P1N1) in between P^0 trials. To assess single-trial learning (column 2) during exposure to the environments, channel trials were randomly interspersed before and after the first P$^+$ trial in a subset of the force-field cycles. Columns 3 to 5 show the output and internal inferences of the COIN model in the same format as Fig. 4c (same parameters as in Fig. 4; Extended Data Fig. 3, parameter set **S**, **E** & **M**). The COIN model qualitatively reproduced the pattern of changes in single-trial learning seen over repeated cycles in this paradigm. As in Fig. 4, differences in the apparent learning rate were not driven by differences in either the proper learning rate (Kalman gain) or the underlying state (column 4) but were instead driven by changes in contextual inference (column 5).



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Cognitive processes and the COIN model. a-**d**, **Maintenance of context probabilities may require working memory. a**, Adaptation in a spontaneous recovery paradigm in which a non-memory (pink) or working memory task (green) is performed at the end of the P^{-} phase before starting the channel-trial phase (data reproduced from Ref. 22). Initial adaptation in the channel-trial phase (inset) shows the working memory task abolishes spontaneous recovery and leads to adaptation akin to evoked recovery (cf. Extended Data Fig. 6a-c). **b**-**d**, COIN model simulation in which the working memory task abolishes the (working) memory of the context responsibilities on the last trial of the P^{-} phase but not the context transition (and thus stationary) probabilities (same parameters as in Fig. 2b and d; Extended Data Fig. 3, parameter set **A**), plotted as in Fig. 2b-c. The circles on the predicted probability (zoomed view) show the values on the first trial in the channel-trial phase. **d**, as (**c**) but for the working memory task. The predicted probabilities on the first trial in the channel-trial phase are the values under the stationary distribution (shown on every trial in the simulation of Extended Data Fig. 1c). We calculate the stationary context distribution by solving $\boldsymbol{\psi} = \boldsymbol{\psi}\hat{\boldsymbol{\Pi}}$ for $\boldsymbol{\psi}$ (a row vector) subject to the constraint that $\boldsymbol{\psi}$ is a valid probability distribution (i.e. all elements of $\boldsymbol{\psi}$ are non-negative and sum to 1), where $\hat{\boldsymbol{\Pi}}$ is the expected local transition probability matrix. **e**-**l**, **Explicit versus implicit learning in the COIN model. e,** Results of a spontaneous recovery paradigm (as in Fig. 2b) for visuomotor learning. Adaptation is computed as participants' reach angle normalised by the size of the experimentally imposed visuomotor rotation. Explicit learning (dark green) is measured by participants indicating their intended reach direction. Implicit learning (light green) is obtained as the difference between total adaptation (solid pink) and explicit learning. In the visual error-clamp phase (P^{c}), participants were told to stop using any aiming strategy so that the direction they moved was taken as the implicit component of learning. A control experiment (dashed pink) was also performed in which there was no reporting of intended reach direction. Data reproduced from Ref. 24. **f**-**l**, Simulation of the COIN model on the same paradigm (same parameters as in Fig. 2b and d but with the addition of a bias parameter; Extended Data Fig. 3, parameter set **A**). **b,** Predictions for experimentally observable quantities. Light green line: implicit learning is the average bias across contexts weighted by the predicted probabilities (cyan line in **j**). Dark green line: explicit learning is the state of the most responsible context on the previous trial (black line in **h**). Solid pink line: total adaptation for the reporting condition is the sum of explicit and implicit learning (as in experiments). Dashed pink line: total adaptation for the non-reporting condition is the average predicted state feedback across contexts weighted by the predicted probabilities (cyan line in **l**, as in all experiments that had no reporting element). **g**-**h**, Inferred bias (**g**) and predicted state (**h**) distributions for each context (colours), with black line showing the mean state of the most responsible context (coloured line below axis) for trials on which an explicit report was solicited. **i**, Predicted probability of each context. Colours as in **g**-**h**, grey is novel context as in Fig. 1f. **j**-**k**, Inferred bias (**j**) and predicted state (**k**) distributions (purple), obtained as mixtures of the respective distributions of individual contexts (**g**-**h**) weighted by their predicted probabilities (**i**), and their means (cyan lines). **l**, Predicted state feedback distribution (purple, computed as the the sum of bias, **j**, and predicted state, **k**) and its mean (cyan).

	single-context models		multiple-context models				
	dual-rate Smith et al.[1]	memory of errors Herzfeld et al.[4]	source of errors Berniker & Körding[16]	winner-take-all Oh & Schweighofer[11]	DP-KF Gershman et al.[12]	MOSAIC Haruno et al.[26]	COIN
spontaneous recovery	✓	✗[a]	✗[b]	✗[b]	✗[c]	✗[d]	✓
evoked recovery	✗[e]	✗[e]	✗[f]	✗[f]	✗[f]	✗[d]	✓
memory updating	✗[g]	✗[g]	✗[g]	✗[h]	✗[g,h]	✓	✓
savings after full washout	✗[i]	✓	✓	✓	✓	✓	✓
anterograde interference	✓	✗[a]	✗[b]	✗[b]	✓	✗[j]	✓
environmental consistency	✗[i]	✓	✗[b]	✗[b]	✗[k]	✓	✓
explicit/implicit learning	✗[m]	✗[l]	✗[l]	✗[l]	✗[l]	✗[l]	✓

Extended Data Table 1 | Comparison of the COIN model to other models. Table shows which experimental phenomena (rows) can be explained by different single and multiple-context models (columns). Alphabetical superscripts index the key feature(s) missing from each model which are primarily responsible for their inability to explain a particular phenomenon. Note that we consider each model as described and implemented by its authors (although it might be possible to modify or extend these models to explain more features). Orange cross-ticks are for models that can partially explain a phenomenon.

Spontaneous recovery, the gradual re-expression of P^+ in the channel-trial phase (Fig. 2c), requires a single-context model to have multiple states that decay on different time scales or a multiple-context model that can change the expression of memories in a gradual manner based on the amount of experience with each context. Therefore, single-context models that have a single state[a], or multiple-context models that do not learn context transition probabilities[b] or do not have state dynamics[d] do not show spontaneous recovery. Models that learn transition probabilities but that do not represent uncertainty about the previous context[c] (the 'local' approximation in DP-KF) can either include a self-transition bias or not. With a self-transition bias, the expression of memories changes in an abrupt manner (akin to evoked recovery) when, in the channel-trial phase, the belief about the previous context changes (e.g. from P^- to P^+), and thus such models fail to explain the gradual nature of spontaneous recovery. Without a self-transition bias, the change in expression of memories is gradual based on updated context counts, but this occurs too slowly relative to the time scale on which the rise of spontaneous recovery occurs.

Evoked recovery, the rapid re-expression of the memory of P^+ in the channel-trial phase (Fig. 2e) that does not simply decay exponentially to baseline (Extended Data Fig. 6e), requires a model to be able to switch between different memories based on state feedback. Therefore, single-context models[e] that cannot switch between memories are unable to show the evoked recovery pattern seen in the data. Multiple-context models with memories that decay exponentially to zero in the absence of observations[f] (as during channel trials) can only partially explain evoked recovery, showing the initial evocation but not the subsequent change in adaptation over the channel-trial phase. Models with no state decay[d] cannot explain evoked recovery.

Memory updating requires a model to update memories in a graded fashion and to use sensory cues to compute these graded updates. Therefore, models that either have no concept of sensory cues[g] or multiple-context models that only update the state of the most probable context in an all-or-none manner[h] do not show graded memory updating.

Savings, faster learning during re-exposure compared to initial exposure, after full washout requires a single-context model to increase its learning rate or a multiple-context model to protect its memories from washout and/or learn context transition probabilities. Therefore, single-context models with fixed learning rates[i] do not show savings.

Anterograde interference, increasing exposure to P^+ leads to slower subsequent adaptation to P^-, requires a single-context model to learn on multiple time scales or a multiple-context model to learn transition probabilities that generalise across contexts. Therefore, single-context models with a single state[a], or multiple-context models that either do not learn transition probabilities[b] or that learn local transition probabilities independently for each row of the transition probability matrix[j] do not show anterograde interference.

Environmental consistency, the increase/decrease in single-trial learning for slowly/rapidly switching environments, requires a model to either adapt its learning rate or learn local transition probabilities based on context transition counts. Therefore, single-context models with fixed learning rates[i] or multiple-context models that either do not learn transition probabilities[b] or that learn non-local transition probabilities based only on context counts[k] do not show the effects of environmental consistency on single-trial learning.

Explicit and implicit learning, the decomposition of visuomotor learning into explicit and implicit components, requires a model to have elements that can be mapped onto these components. For most models, there is no clear way to map model elements onto these components[l]. It has been suggested that the fast and slow processes of the dual-rate model correspond to the explicit and implicit components of learning, respectively. However, in a spontaneous recovery paradigm, this mapping only holds during initial exposure and fails to account for the time course of the implicit component during the counter-exposure and channel-trial phases[m] (see Suppl. Inf.).

Supplementary Information

1 Experimental methods

1.1 Participants

A total of 40 neurologically-healthy participants (18 males and 22 females; age 27.7 ± 5.6 yr, mean ± s.d.) were recruited to participate in two experiments, which had been approved by the Cambridge Psychology Research Ethics Committee and the Columbia University IRB (AAAR9148). All participants provided written informed consent and were right-handed according to the Edinburgh handedness inventory[1]. To provide sufficient power, sample sizes were chosen on the basis of the typical between-participant variability observed in similar motor adaptation studies[2–5].

1.2 Experimental apparatus and approach

All experiments were performed using a vBOT planar robotic manipulandum with virtual-reality system and air table[6]. The vBOT is a modular, general-purpose, two-dimensional planar manipulandum optimised for dynamic learning paradigms. The vBOT's handle position was measured using optical encoders sampled at 1 kHz while torque motors allowed forces to be generated at the handle and updated at the same rate. Participants grasped the handle of the manipulandum with their right hand while their forearm was supported on an air sled, which constrained arm movements to the horizontal plane and reduced friction.

A monitor mounted horizontally face-down above the vBOT projected images via a horizontal mirror so that visual feedback was overlaid in the plane of movement. In the spontaneous/evoked recovery experiment, the mirror prevented direct vision of the hand and forearm. In the memory updating experiment, a semi-silvered mirror was used and a lamp illuminated the hand from below the mirror with the illumination adjusted so that both the vBOT, hand, arm and virtual images were clearly visible. This was done to ensure that participants had an accurate estimate of the state of their hand and arm (as in Ref. 4).

The manipulandum controlled a virtual "object" (cursor or rectangular tool, depending on the experiment) that was displayed centred on the hand and translated with hand movements (Fig. 2a & Fig. 3a). On each trial, participants first aligned the centre of the object with the home position (0.5 cm radius circle) situated in the midline approximately 30 cm in front of the participant's chest. The trial started after the centre of the object was within 0.5 cm of the home position and had remained below a speed of 0.5 cm·s^{-1} for 0.1 s. After a 0.3 s delay, a target (a circle with a radius of 0.5 cm) appeared 12 cm away (distally within the sagittal plane), with the transverse position depending on the experiment (see below). A tone indicated that the participants should initiate a reaching movement to the target. Participants were instructed to move the object (or a specific control point on it, depending on the experiment, see below) to the target. In all cases, the shortest hand movement path connected the centre of the object to the target in a straight line within the sagittal plane. The trial ended when the control point had remained within 0.5 cm of the target for 0.1 s below a speed of 0.5 cm·s^{-1}. If the peak speed of the movement was less than 50 cm·s^{-1} or more than 70 cm·s^{-1}, a low-pitch tone sounded and a 'too slow' or 'too fast' message was displayed, respectively. At the end of each trial, the vBOT actively returned the hand to the home position.

On each trial, the vBOT could either generate no forces (P^0, null field), a velocity-dependent curl force field (P$^+$ or P$^-$ perturbation depending on the direction of the field) or a force channel (P^c, channel trials).

For the curl force field, the force generated on the hand was given by

$$\begin{bmatrix} F_\mathrm{x} \\ F_\mathrm{y} \end{bmatrix} = g \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \tag{S1}$$

where F_x, F_y, $\dot{x}$ and $\dot{y}$ are the forces and velocities at the handle in the x (transverse) and y (sagittal) directions, respectively. The gain g was set to ±15 N·s·m^{-1}, where the sign of g specified the direction of the curl field (counterclockwise or clockwise which were assigned to P^{+} and P^{-}, counterbalanced across participants). On channel trials, the hand was constrained to move along a straight line from the home position to the target. This was achieved by simulating forces associated with a stiff spring and damper, with the forces acting perpendicular to the long axis of the channel. A spring constant of 3,000 N·m^{-1} and a damping coefficient of 140 N·s·m^{-1} were used. Channel trials clamped the kinematic error close to zero and were used to measure the participant's level of adaptation to the P^{+} and P^{-} perturbations based on the forces they generated into the channel walls[7,8].

1.3 Experiment 1: spontaneous and evoked recovery

Participants either performed a spontaneous (n=8) or evoked (n=8) recovery condition. In both conditions, the virtual object controlled by participants was simply a cursor (blue 0.4 cm radius disc), which was always aligned with the centre of the handle. The control point was the centre of the cursor (unmarked).

1.3.1 Spontaneous recovery condition

In the spontaneous recovery condition, participants (5 males and 3 females; age 32.1 ± 7.1 yr, mean ± s.d.) performed a version of the standard spontaneous recovery paradigm[3]. The paradigm consisted of a pre-exposure phase (5 blocks, with 10 trials each) with a null field (P^{0}). This was followed by an exposure phase (12 blocks, with 10 trials each, and an additional 5 exposure trials after the 45 s rest break given after block 6) with P^{+} (the direction of the force field assigned to P^{+} was counterbalanced across participants). In the pre-exposure and exposure phases, to assess adaptation, each block of 10 trials had one channel trial (P^{c}) in a random location (not the first). After the exposure phase, participants were rapidly de-adapted in a counter-exposure phase by applying 15 trials with the opposite perturbation (P^{-}). This was followed by a long series of 150 channel trials (P^{c}).

1.3.2 Evoked recovery condition

In the evoked recovery condition, participants (3 males and 5 females; age 27.2 ± 5.9 yr, mean ± s.d.) performed a modified version of the spontaneous recovery paradigm which differed in that the 3rd and 4th trials of the channel-trial phase were replaced with P^{+} trials (Fig. 2d).

1.4 Experiment 2: memory updating

This experiment was based on a paradigm in which sensory cues allow multiple memories to be learned simultaneously[4] and involved n=24 participants (10 males and 14 females; age 26.4 ± 4.2 yr, mean ± s.d.). The virtual object controlled by participants was a solid green rectangle (16×3 cm, width×depth, with a yellow cross indicating its centre) was displayed centred on the hand (Fig. 3a). The object also had two control points (blue 0.4 cm radius discs) ± 7 cm lateral to the centre of the object. Targets were in front of either the left or right control point. If the target was aligned with the left control point, participants were instructed to move the left control point to the target, and conversely for the target aligned with the right control point. Crucially, because each target was aligned with its respective control point, the hand had to move to the same location to attain either target. The different targets required participants to attend to either of the two control points and thus provided distinctive sensory cues for the trial[4]. We indicate the sensory cue used on a trial by a subscript (e.g. P_1^+ and P_2^+ for the P^+ perturbation with the left and right sensory cue, respectively). The experiment consisted of three phases: pre-training, training and post-training. The training phase (see details below) consisted of exposure to two cue-perturbation pairs (P_1^+ and P_2^-) that differed in both the perturbation and the sensory cue so that participants could be expected to associate each cue with its corresponding perturbation.

In the pre-training and post-training phases (Fig. 3c) participants performed blocks of trials which consisted of a variable number of P^0 washout trials (8, 10 or 12 in the pre-training phase and 2, 4 or 6 in the post-training phase) with an equal number of each sensory cue in a pseudorandom order, followed by a triplet of trials to assess single-trial learning (see below). The P^0 trials were used to bring adaptation close to baseline before the triplet of trials. The first and third trial in the triplet were always channel trials with sensory cue 1 (P_1^c) and the middle trial of the triplet ('exposure' trial) was one of four possible combinations of perturbation sign (force-field direction) and sensory cue (control point): P_1^+, P_2^-, P_2^+ and P_1^- (Fig. 3c). Therefore, the first two exposure trial types (P_1^+ and P_2^-) were the same as those experienced in the training phase and thus in the post-training phase provided consistent evidence about the contexts experienced during the training phase, whereas the latter two (P_2^+ and P_1^-) were different from those experienced in the training phase and thus provided conflicting evidence about the contexts experienced during the training phase. Within each sequence of 4 blocks, each of these combinations was experienced once and the four blocks were repeated 4 times in pre-training and 8 times in post-training. Importantly, the relationship between sensory cues and perturbations was balanced, such that each triplet type was presented an equal number of times and each cue was presented an equal number of times in the P^0 trials.

In the training phase (Fig. 3b), each sensory cue was consistently and repeatedly associated with one perturbation (P_1^+ and P_2^-) during force-field trials, with additional channel trials before and after these trials to assess how learning progressed, as well as occasional channel triplets (using consistent exposure trials only) to assess single-trial learning (preceded by washout trials, as explained above). To do this, participants performed 24 blocks, each consisting of 62-70 trials presented in the following order:

- 2 channel trials (one P_1^c and P_2^c, order counterbalanced across consecutive blocks);
- 32 force-field trials (equal number of P_1^+ and P_2^- within each 8 trials in a pseudorandom order);
- 2 channel trials (one P_1^c and P_2^c order counterbalanced across consecutive blocks);
- 14, 16 or 18 washout trials (equal number of P_1^0 and P_2^0 in a pseudorandom order);
- 1 triplet (exposure trial of P_1^+ or P_2^- counterbalanced across consecutive blocks);
- 6, 8 or 10 washout trials (equal number of P_1^0 and P_2^0 in a pseudorandom order);
- 1 triplet (exposure trial of P_1^+ or P_2^-, whichever was not used on the previous triplet).

We sampled without replacement the number of null-field trials from the options above and replenished these options whenever they emptied.

A 60 s rest break was given after every 3 blocks during the training phase. After each rest break, 8 null-field trials were performed in which the sensory cues were presented in a pseudorandom order.

The control point assigned to sensory cue 1 (used on all triplet channel trials) and sensory cue 2 was counterbalanced across participants as was the direction of the force field assigned to P$^+$ and P$^-$.

Prior to the experiment, participants performed a familiarisation phase of 80 trials consisting of null-field trials and channel trials for each sensory cue in a pseudorandom order.

1.5 Data analysis

On channel trials, we calculated adaptation as the proportion of the force field that was compensated for by the participant. This was taken as the slope of the regression (with zero offset) of the time series of actual (signed) force generated into the channel walls against the time series of forces (based on the hand velocity in the channel) that would fully compensate for the perturbation had it been present[3]. For this analysis, we used the portion of the movement where the hand velocity was greater than 1 cm·s^{-1}. Single-trial learning was calculated as the change in adaptation between the first and second channel trial of a triplet (Fig. 3c).

To identify changes in single-trial learning between triplets in the memory updating experiment, two-way repeated-measures ANOVAs were performed with factors of cue (2 levels: cue 1 and cue 2) and perturbation (2 levels: P$^+$ and P$^-$). To test whether the modulatory effects of cue and perturbation were confined to separate subsets of participants, we quantified the effect of each by computing, on an individual-participant basis, the following contrasts in single-trial learning: P$^+_1$ + P$^-_1$ − P$^+_2$ − P$^-_2$ (cue effect) and P$^+_1$ + P$^+_2$ − P$^-_1$ − P$^-_2$ (perturbation effect). These are the same contrasts that underlie the ANOVA-based analysis we conducted to test whether each manipulation had an overall effect across participants. We then split participants into 2×2 groups based on whether each effect was below or above a threshold-level. If separate subsets of participants showed each effect, a Fisher's exact test should indicate a significant difference between the resulting 2×2 histogram and its surrogate that assumes that the two binarised effects are distributed independently across participants. The non-significant results reported in the main text used a median split for both effects (0.08 for cue, and 0.27 for perturbation; Extended Data Fig. 7f). We obtained essentially identical results when we instead used a split at 0 for both effects (odds ratio $= 1.3$, $p = 1.00$). All statistical tests were two-sided. Data analysis was performed using MATLAB R2020a.

2 COIN model

2.1 Generative model

The generative model that underlies the COIN model is described in the Methods. Here we first give details of a stick-breaking representation of the distributions of the infinite global and local transition (or cue) probability vectors under the hierarchical Dirichlet process (HDP, a key component of the generative model). We then present an alternative representation of the HDP known as the Chinese restaurant franchise, which allows a sequence of contexts (or cues) to be sampled directly from the prior of the HDP by marginalising out the infinite global and local transition (or cue) probabilities, such that they never need to be explicitly represented. In turn, this representation allows efficient posterior inference algorithms, which we describe in Section 2.3.

2.1.1 The stick-breaking construction

The infinite global transition probability vector $\boldsymbol{\beta} = (\beta_j)_{j=1}^{\infty}$ obeys a $\mathrm{GEM}(\gamma)$ distribution, which can be sampled from via a 'stick-breaking' construction that is analogous to recursively breaking a stick of length 1 into infinitely many pieces:

$$\beta_j = \beta'_j \prod_{i=1}^{j-1} (1 - \beta'_i) \qquad \beta'_j \mid \gamma \sim \mathrm{Beta}(1, \gamma)\,. \tag{S2}$$

In each step $j = 1, \ldots, \infty$, a portion $\beta'_j \in [0, 1]$ of the remaining stick, which has length $\prod_{i=1}^{j-1} (1 - \beta'_i)$, is broken off and assigned to β_j. This guarantees that $0 \leq \beta_j \leq 1$ and $\sum_{j=1}^{\infty} \beta_j = 1$, as required for a set of probabilities. The probabilities generated by this stick-breaking construction decay exponentially as a function of j in expectation, with the hyperparameter γ controlling the rate of decay:

$$\mathbb{E}[\beta_j] = \frac{1}{1+\gamma} \left(\frac{\gamma}{1+\gamma} \right)^{j-1}. \tag{S3}$$

For context j, the infinite local transition probability vector $\boldsymbol{\pi}_j = (\pi_{jk})_{k=1}^{\infty}$ obeys a Dirichlet process distribution $\mathrm{DP}\left(\alpha + \kappa, \frac{\alpha\,\boldsymbol{\beta} + \kappa\,\boldsymbol{\delta}_j}{\alpha+\kappa}\right)$, which can be sampled from by drawing an infinite set of stick-breaking weights $\tilde{\boldsymbol{\pi}}_j = (\tilde{\pi}_{jk})_{k=1}^{\infty}$ via a stick-breaking construction, associating each weight with a 'to context' by drawing a corresponding infinite set of variables $(\tilde{\chi}_{jk})_{k=1}^{\infty}$ from a discrete distribution (each $\tilde{\chi}_{jk} \in \{1, \ldots, \infty\}$ represents the identity of the 'to context' associated with weight $\tilde{\pi}_{jk}$) and summing weights that are associated with the same 'to context':

$$\boldsymbol{\pi}_j = \sum_{k=1}^{\infty} \tilde{\pi}_{jk}\, \boldsymbol{\delta}_{\tilde{\chi}_{jk}} \qquad \tilde{\chi}_{jk} \mid \alpha, \boldsymbol{\beta}, \kappa \sim \mathrm{Discrete}\left(\frac{\alpha\,\boldsymbol{\beta} + \kappa\,\boldsymbol{\delta}_j}{\alpha+\kappa} \right) \qquad \tilde{\boldsymbol{\pi}}_j \mid \alpha, \kappa \sim \mathrm{GEM}(\alpha+\kappa)\,. \tag{S4}$$

Here $\boldsymbol{\delta}_{\tilde{\chi}_{jk}}$ is an infinite-dimensional one-hot vector with the $\tilde{\chi}_{jk}^{\text{th}}$ element set to 1 and all other elements set to 0. Note that this is a two-level hierarchical process; global transition probabilities generated at the top level (Eq. S2) are used to generate local transition probabilities at the bottom level (Eq. S4).

Analogous constructions can be used to sample the infinite global and local cue probability vectors but with γ replaced with γ^{e}, α replaced with α^{e} and κ (the self-transition bias parameter) set to zero.

2.1.2 The Chinese restaurant franchise with loyal customers

Here we present an alternative representation of the hierarchical Dirichlet process that provides a mechanism for sampling sequences of contexts and cues from the prior of the COIN model as well as a framework for posterior inference. In addition, this representation provides intuitions for how the generative process works as trials are experienced.

In the Chinese restaurant franchise (CRF)[9], there are an infinite number of restaurants each with an infinite number of tables. Each table serves only one dish from an infinite global menu shared by all the restaurants (hence the franchise). The same dish can be served on multiple tables in the same restaurant as well as in multiple restaurants. Each customer enters a restaurant and is seated at a table where a dish is served.

In the COIN model, customers correspond to trials and will arrive in the same temporal order as the trials. For both context transitions and cue emissions, the restaurant that the customer enters corresponds to the current context (i.e. if context t is $c_t = j$, customer t enters restaurant j). For context transitions, the dish served at the table at which the customer sits corresponds to the next context (i.e. if dish k is served, a transition to context $c_{t+1} = k$ occurs). For cue emissions, the dish served at the table corresponds to the sensory cue emitted on that trial (i.e. if dish k is served, cue $q_t = k$ is emitted). Note that separate CRFs are used for context transitions and cue emissions.

Although there an infinite number of restaurants, tables and dishes in the franchise, to generate a finite amount of data, we only need to consider the finite number of occupied tables and the finite number of dishes served at those tables (i.e. the contexts and sensory cues already experienced), as well as one empty table in each occupied restaurant and one novel dish (so that a novel context or sensory cue can be experienced). See also note in Section 2.2.

Table assignment and dish selection

Let $c_{1:t-1}$, $\tau_{1:t-1}$, $k_{1:t-1}$ be the sequence of restaurants, tables and dishes associated with the first $t-1$ customers. Let us define the following summary statistics of these past customers: J is the number of restaurants with at least one customer, K is the number of unique dishes across the franchise served to at least one customer (for the context-CRF $K = J$, as we will see below), the elements of the $J \times K$ matrix $\mathbf{M}$, m_{jk}, store the number of tables in restaurant j already serving dish k, with sums across columns, $\mathbf{m}^{\mathrm{t}} = \mathbf{M}\,\mathbf{1}$, and rows, $\mathbf{m}^{\mathrm{d}} = \mathbf{M}^{\mathsf{T}}\,\mathbf{1}$, respectively counting the number of occupied tables in each restaurant and the number of tables serving each dish across the franchise, and $\{\tilde{\mathbf{n}}_j\}_{j=1}^{J}$ is a set of vectors with elements $\tilde{n}_{j\tau}$ counting the number of customers in restaurant j sitting at table $\tau = 1 \ldots m_j^{\mathrm{t}}$ so far.

Let us assume that customer t was allocated to restaurant $c_t = j$ (the way this allocation is made will be described below). Then the table chosen by customer t in that restaurant, τ_t, is randomly sampled as

$$\tau_t \mid \tau_{1:t-1}, c_{1:t-1}, c_t = j, \alpha \sim \mathrm{Discrete}\left(\frac{\left[\tilde{\mathbf{n}}_j^{\mathsf{T}},\, \alpha\right]}{\tilde{\mathbf{n}}_j^{\mathsf{T}}\,\mathbf{1} + \alpha}\right). \tag{S5}$$

Thus the customer either sits at an occupied table, $1 \leq \tau_t \leq m_j^{\mathrm{t}}$, with probability proportional to the number of people already sitting at that table (and $\tilde{n}_{j\tau_t} \leftarrow \tilde{n}_{j\tau_t} + 1$), or sits at a new table, $\tau_t = m_j^{\mathrm{t}} + 1$, with probability proportional to α (and in this case $\tilde{\mathbf{n}}_j \leftarrow \left[\tilde{\mathbf{n}}_j^{\mathsf{T}},\, 1\right]^{\mathsf{T}}$). The hyperparameter α controls how the number of occupied tables grows as a function of the number of customers in the restaurant. With small α, most customers will sit at the same table, and so the number of occupied tables will grow slowly over trials. This table assignment process has the effect that tables with many customers attract even more customers.

The dish served to customer t, k_t, is the same as that served to all previous customers sitting at the same table. Otherwise, if this customer is the first to sit at this table, the dish for the table is randomly sampled as

$$k_t \mid \tau_{1:t-1}, k_{1:t-1}, c_{1:t-1}, c_t = j, \gamma \sim \mathrm{Discrete}\left(\frac{\left[{\mathbf{m}^{\mathrm{d}}}^{\mathsf{T}},\, \gamma\right]}{{\mathbf{m}^{\mathrm{d}}}^{\mathsf{T}}\,\mathbf{1} + \gamma}\right). \tag{S6}$$

Thus when the customer sits at a new table, they are either served an existing dish, $1 \leq k_t \leq K$, with probability proportional to the number of tables already serving that dish in the franchise (and $m_{jk_t} \leftarrow m_{jk_t} + 1$), or they are served a new dish, $k_t = K + 1$, with probability proportional to γ (and in this case $\mathbf{M} \leftarrow \left[\mathbf{M},\, \boldsymbol{\delta}_j\right]$ and $K \leftarrow K+1$). The hyperparameter γ is analogous to that of α for table assignments: it controls how the number of dishes grows as a function of the number of tables. With small γ, most tables will have the same dish, and so the number of dishes (i.e. contexts and cues) will grow slowly over trials. With large γ, most tables will have different dishes, and so the number of dishes (i.e. contexts and cues) will grow rapidly over trials. As a result, this dish assignment process is similar to that for tables: it has the effect that dishes already served at many tables in the franchise will be served at even more tables, and since the dish served to the customer ultimately determines the next context (or current cue), this makes commonly experienced transitions (or cues) increasingly likely in the future.

Note that although separate CRFs are used for context transitions and cue emissions with respect to table assignment and dish selection, the CRFs are not independent, as the restaurant to which customer t is allocated, c_t, is decided by the dish served to the previous customer in the CRF for context transitions:

$$c_t = k_{t-1}^{\text{context}}, \tag{S7}$$

(and if this is the first customer to enter this restaurant, $c_t > J$, then $\mathbf{M} \leftarrow \left[\mathbf{M}^{\mathsf{T}}, \mathbf{0}\right]^{\mathsf{T}}$ and $J \leftarrow J + 1$).

This completes the descriptions of these two CRFs as we have fully defined how $c_{1:t-1}$, $\tau_{1:t-1}$, $k_{1:t-1}$ determines c_t (Eq. S7), τ_t (Eq. S5) and k_t (Eq. S6).

Loyal customers

For the context transitions, the process we have described so far has no self-transition bias. To include such a bias (as in the COIN model), the CRF can be extended to include loyal customers[10].

Each restaurant now has a specialty dish whose index is the same as that of the restaurant (e.g. dish j is the specialty dish of restaurant j). The specialty dish is available in all restaurants, but is more popular in the dish's namesake restaurant. This leads to family loyalty to a restaurant, as the increased popularity of the specialty dish means that children are more likely to eat at the same restaurant as their parent. Hence, multiple consecutive generations often eat at the same restaurant.

To simplify inference in the CRF with loyal customers, a distinction is made between a *considered* dish, $\bar{k}_t$, and a *served* dish, k_t. This also requires us to introduce analogous additional summary statistics of past customers: $\bar{K}$ is the number of unique dishes across the franchise considered by at least one customer, the elements of the $J \times \bar{K}$ matrix $\bar{\mathbf{M}}$, $\bar{m}_{jk}$, store the number of tables in restaurant j at which dish k was considered, with sums across columns, $\bar{\mathbf{m}}^{\mathrm{t}}$, and rows, $\bar{\mathbf{m}}^{\mathrm{d}}$, as before. In addition, a new parameter κ controls the strength of the loyalty effect.

We again assume that customer t was allocated to restaurant $c_t = j$ (the way this allocation is done is unchanged from Eq. S7). Then the table chosen by customer t in that restaurant is determined analogously to the previous setup (Eq. S5) with α replaced by $\alpha + \kappa$:

$$\tau_t \mid \tau_{1:t-1}, c_{1:t-1}, c_t = j, \alpha, \kappa \sim \text{Discrete}\left(\frac{\left[\tilde{\mathbf{n}}_j^{\mathsf{T}}, \alpha + \kappa\right]}{\tilde{\mathbf{n}}_j^{\mathsf{T}}\,\mathbf{1} + \alpha + \kappa}\right). \tag{S8}$$

As before, a customer choosing an already occupied table eats the dish that is already served there. Otherwise, the first customer to sit at a table considers a dish for that table without acknowledging the increased popularity of the specialty dish of the restaurant, i.e. analogously to how dishes were served in the previous setup (Eq. S6, but depending on the popularity of previously considered dishes, rather than previously served dishes):

$$\bar{k}_t \mid \tau_{1:t-1}, \bar{k}_{1:t-1}, c_{1:t-1}, c_t = j, \gamma \sim \text{Discrete}\left(\frac{\left[\bar{\mathbf{m}}^{\mathrm{d}\,\mathsf{T}}, \gamma\right]}{\bar{\mathbf{m}}^{\mathrm{d}\,\mathsf{T}}\,\mathbf{1} + \gamma}\right). \tag{S9}$$

However, with some probability $\rho = \kappa/(\alpha + \kappa)$ (which acts as a normalised self-transition bias), this considered dish is overridden (perhaps by a waiter's suggestion) and the specialty dish is served instead:

$$k_t \mid \bar{k}_t, c_t = j, \alpha, \kappa \sim \text{Discrete}\big((1 - \rho)\ \boldsymbol{\delta}_{\bar{k}_t} + \rho\,\boldsymbol{\delta}_j\big). \tag{S10}$$

The distribution of served dishes in the CRF with loyal customers can be related to the global transition probabilities β of the stick-breaking representation (Section 2.1.1). This relationship allows the CRF to be used to infer the global transition (and cue) probabilities at inference (Sections 2.3.7 and 2.3.8). Each dish on the infinite global menu of dishes has an overall popularity or rating that determines the distribution of dishes in each restaurant. In the special case when $\rho = 0$, the served dishes are distributed as $\boldsymbol{\beta}$, regardless of the restaurant. In contrast, when $\rho > 0$, the increased popularity of the specialty dish leads to modified dish ratings, with the served dishes in restaurant j being distributed as $(\alpha\, \boldsymbol{\beta} + \kappa\, \boldsymbol{\delta}_j)/(\alpha + \kappa)$. These modified dish ratings correspond to the expected local transition probabilities under the Dirichlet process prior (Eq. 8).

2.2 A note on the hypothesis space of contextual inference

To perform exact contextual inference, all possible context sequences should be considered, with each sequence assigned a posterior probability. However, in practice, this is infeasible, as the number of possible context sequences grows rapidly over time. For example, in an environment with C contexts, the number of possible context sequences over t time steps is C^t. This is a vast hypothesis space even at a moderate number of time points (consider just two contexts and 50 time steps). Furthermore, if the number of contexts in the environment is unknown and unbounded (as in the COIN model), the number of contexts that need to be considered in a sequence grows with the length of the sequence, as a novel context could have become active at each point in time. To deal with this complexity, rather than considering all possible context sequences, a smaller, tractable subset of context sequences can be considered instead. This is the strategy employed by the inference algorithm of the COIN model, which uses particles to sample context sequences according to their posterior probability. If the probability of a particular context sequence is small under the exact posterior, a proportionately small fraction of particles (or perhaps even none given that a finite number of particles are used in practice) will sample this sequence.

2.3 Inference with particle learning

The goal of inference is to estimate a joint posterior distribution over the number of contexts, the current context (e.g. in Fig. 1c, the identity of the currently manipulated object, such as a cup or a sugar bowl), the current state of each context (e.g. the current weight of the cup) and the parameters governing the state dynamics (e.g. how quickly liquid empties when the cup is tilted), the context transitions (e.g. that we tend to handle the sugar bowl once we have filled our cup) and the cue emissions (e.g. that cups tend to have a similar visual appearance) at each point in time based on the state feedback (e.g. the noisy weight of the currently manipulated object, purple dots) and sensory cue (e.g. visual appearance of the currently manipulated object, green and yellow background colour) observations made so far. To perform posterior inference in an online (i.e. recursive) manner, we use a sequential Monte Carlo (simulation-based) method known as particle learning[11,12].

Particle learning extends standard particle filtering methods by incorporating the estimation of time-invariant parameters via a fully-adapted filter that utilises conditional sufficient statistics for the parameters. To sequentially compute a particle approximation to the joint posterior distribution of contexts, states and conditional sufficient statistics for the parameters, an essential state vector is constructed and is used together with a predictive distribution and propagation rule to build a resampling-sampling framework.

Central to particle learning is the essential state vector z_t that contains samples and/or sufficient statistics of the contexts, states and parameters. Online context and state filtering and parameter learning is equivalent to sequential filtering of the essential state vector:

$$p(z_t|\mathcal{D}_{1:t}) \propto \int p(z_t|z_{t-1}, \mathcal{D}_t)p(\mathcal{D}_t|z_{t-1})p(z_{t-1}|\mathcal{D}_{1:t-1})\mathrm{d}z_{t-1}, \tag{S11}$$

where $\mathcal{D}_{1:t} = \{\mathcal{D}_1, \ldots, \mathcal{D}_t\}$ is the sequence of observations.

Particle learning uses an ensemble of particles $\mathcal{Z}_t = \{z_t^{(i)}\}_{i=1}^P$ that are equally weighted to form a discrete approximation to the filtering distribution $p(z_t|\mathcal{D}_{1:t})$ via

$$\hat{p}(z_t|\mathcal{D}_{1:t}) = \frac{1}{P}\sum_{i=1}^{P}\delta(z_t - z_t^{(i)}), \tag{S12}$$

where $\delta(\cdot)$ is the Dirac delta function. A recursive formula for obtaining $\hat{p}(z_t|\mathcal{D}_{1:t})$ from $\hat{p}(z_{t-1}|\mathcal{D}_{1:t-1})$ is suggested by the decomposition shown in Eq. S11. First, in a *resample* step, particles are sampled with replacement from a multinomial distribution with weights proportional to the predictive distribution $p(\mathcal{D}_t|z_{t-1})$. This produces a particle approximation to the smoothed distribution $p(z_{t-1}|\mathcal{D}_{1:t})$ by replicating/discarding particles based on how well they predicted the observations at time t. Then, in a *propagate* step, the resampled particles are propagated via the evolution equation $p(z_t|z_{t-1}, \mathcal{D}_t)$. A final *sample* step can also be performed in which new parameters are sampled from their updated posterior distributions conditioned on the propagated essential state vectors. Although this last step is optional, without it the diversity of parameters would reduce with each resampling step until all particles shared the same parameters, a problem known as degeneracy. A single time step of particle learning is summarised in Algorithm 1.

```
resample $\{z_{t-1}^{(i)}\}_{i=1}^P$ with weights $w_t^{(i)} \propto p(\mathcal{D}_t|z_{t-1}^{(i)})$ to obtain $\{\tilde{z}_{t-1}^{(i)}\}_{i=1}^P$ and reset particle weights to $1/P$
for $i = 1, \ldots, P$ do
  propagate $\tilde{z}_{t-1}^{(i)}$ to $z_t^{(i)}$ via $p(z_t|\tilde{z}_{t-1}^{(i)}, \mathcal{D}_t)$
  sample $\theta^{(i)}$ from $p(\theta|z_t^{(i)})$
end for
```

Algorithm 1: The general particle learning algorithm.

In the COIN model, the essential state vector $z_t = \{c_t, s_t^x, s_t^\theta, \theta\}$ contains the context c_t, the sufficient statistics (mean and variance) for the states s_t^x, the sufficient statistics for the parameters s_t^θ and the parameters θ. Following the direct assignment algorithm of Ref. 9, we do not sample the local transition and cue distributions. Instead, we sample the global transition and cue distributions and integrate out the local distributions by computing their expected values (Eqs. S15 and S16). Hence, in the COIN model $\theta = \{\boldsymbol{\beta}, \boldsymbol{\beta}^{\mathrm{e}}, \{\boldsymbol{\omega}^{(j)}\}_j\}$. The propagate step in Algorithm 1 can be decomposed into three separate steps:

$$p(z_t|z_{t-1}, y_t, q_t) = \underbrace{p(s_t^\theta|s_t^x, c_t, z_{t-1}, y_t, q_t)}_{\text{propagate } s_{t-1}^\theta}\underbrace{p(s_t^x|c_t, z_{t-1}, y_t)}_{\text{propagate } s_{t-1}^x}\underbrace{p(c_t|z_{t-1}, y_t, q_t)}_{\text{propagate } c_{t-1}}. \tag{S13}$$

First, the context is propagated conditioned on the state feedback and the sensory cue. Then, the sufficient statistics for the states are propagated conditioned on the context and the state feedback. Finally, the sufficient statistics for the parameters are sampled conditioned on the context, the sufficient statistics for the states, the state feedback and the sensory cue.

We now describe the resample, propagate and sample steps in detail for the COIN model. Note that for simplicity, we suppress the superscript notation that indexes particles except where it is necessary (e.g. when summing over particles, as in Eqs. S29 and S37).

2.3.1 Resample

Given the particle approximation $\hat{p}(z_{t-1}|y_{1:t-1}, q_{1:t-1})$, the updated smoothed approximation $\hat{p}(z_{t-1}|y_{1:t}, q_{1:t})$ is obtained by resampling particles with weights w_t proportional to the predictive distribution:

$$\begin{aligned} w_t &\propto p(y_t, q_t|z_{t-1}) \\ &= \sum_{j=1}^{C+1} p(c_t = j, y_t, q_t|z_{t-1}) \\ &= \sum_{j=1}^{C+1} p(c_t = j|z_{t-1})p(q_t|c_t = j, z_{t-1})p(y_t|c_t = j, z_{t-1}), \end{aligned} \tag{S14}$$

where C is the number of contexts known up to trial $t-1$. The sum over contexts is to $C+1$ to include the possibility that the latest observations were generated by a novel context, the $(C+1)^{\text{th}}$ context.

The first term of the predictive distribution is the expected local transition probability, which can be written as

$$p(c_t|z_{t-1}) = \frac{\alpha\beta_{c_t} + \kappa\delta_{c_{t-1}c_t} + n_{c_{t-1}c_t}}{\alpha + \kappa + n_{c_{t-1}\cdot}}, \tag{S15}$$

where $\delta_{c_{t-1}c_t}$ is the Kronecker delta that is equal to 1 if $c_{t-1} = c_t$ and 0 otherwise and $n_{c_{t-1}c_t}$ denotes the number of transitions from context c_{t-1} to context c_t up to trial $t-1$. Dots represents marginal counts. For example, $n_{c_{t-1}\cdot} = \sum_{j=1}^{C} n_{c_{t-1}j}$ is the number of transitions out of context c_{t-1} up to trial $t-1$. Note that the probability of transitioning to context c_t depends on the global transition probability β_{c_t}, regardless of the identity of the previous context c_{t-1}. Thus when the global transition distribution is updated (see Section 2.3.8), the local transition probabilities from all contexts are also updated (Extended Data Fig. 1b). Importantly, this means that transition probabilities learned in one context generalise to all contexts.

The second term of the predictive distribution is the expected local cue probability, which can be written as

$$p(q_t|c_t, z_{t-1}) = \frac{\alpha^{\text{e}}\beta^{\text{e}}_{q_t} + n^{\text{e}}_{c_tq_t}}{\alpha^{\text{e}} + n^{\text{e}}_{c_t\cdot}}, \tag{S16}$$

where $n^{\text{e}}_{c_tq_t}$ denotes the number of emissions of cue q_t in context c_t up to trial $t-1$, $n^{\text{e}}_{c_t\cdot} = \sum_{q_t=1}^{Q} n^{\text{e}}_{c_tq_t}$ is the number of cues emitted in context c_t up to trial $t-1$ and Q is the number of cues emitted up to trial $t-1$.

The third term of the predictive distribution depends on the predicted state feedback in each context and is given by

$$p(y_t|c_t, z_{t-1}) = \mathcal{N}(\hat{y}^{(c_t)}_t, p^{(c_t)}_t), \tag{S17}$$

where $\hat{y}^{(c_t)}_t$ and $p^{(c_t)}_t$ are the mean and variance of the predicted state feedback distribution for context c_t provided by the time update equations of the Kalman filter (Algorithm 2).

The particles of the smoothed approximation $\hat{p}(z_{t-1}|y_{1:t}, q_{1:t})$ are propagated via the three steps outlined in Eq. S13, which we now describe in Sections 2.3.2 to 2.3.5 in detail.

```
for $j = 1, \ldots, C+1$ do
  if $j \leq C$ then
    $\hat{x}^{(j)}_{t|t-1} = a^{(j)} \hat{x}^{(j)}_{t-1|t-1} + d^{(j)}$
    $v^{(j)}_{t|t-1} = a^{(j)} v^{(j)}_{t-1|t-1} a^{(j)} + \sigma^2_{\mathrm{q}}$
  else if $j = C+1$ then
    $\hat{x}^{(j)}_{t|t-1} = d^{(j)} / (1 - a^{(j)})$
    $v^{(j)}_{t|t-1} = \sigma^2_{\mathrm{q}} / (1 - \left[a^{(j)}\right]^2)$
  end if
  $\hat{y}^{(j)}_t = \hat{x}^{(j)}_{t|t-1}$
  $p^{(j)}_t = v^{(j)}_{t|t-1} + \sigma^2_{\mathrm{r}}$
end for
```

Algorithm 2: State and state feedback prediction. Mean and variance of the predicted state distribution and the predicted state feedback distribution for each known context ($j \leq C$) and a novel context ($j = C+1$). For a novel context, the mean and variance of the predicted state distribution are equal to the mean and variance of the stationary distribution (Eq. 4) conditioned on state retention and drift parameters sampled from the prior.

2.3.2 Propagate the context

The context is propagated by sampling $c_t \in \{1, \ldots, C+1\}$ from the 'responsibilities'

$$p(c_t | z_{t-1}, y_t, q_t) \propto p(c_t, y_t, q_t | z_{t-1}), \tag{S18}$$

where $p(c_t, y_t, q_t | z_{t-1})$ is given in Eq. S14. As can be appreciated from Eq. S14, contextual inference fuses information from multiple sources (Fig. 1b, open arrows). First, it uses prior expectations $p(c_t = j | z_{t-1})$ about which context the learner is in based on the history of contexts inferred so far (the global transition probability of each context as well as the context transition counts). Second, it evaluates the likelihoods that the current sensory cue and state feedback observations are generated by each context, $p(q_t | c_t = j, z_{t-1})$ and $p(y_t | c_t = j, z_{t-1})$, respectively. Note that the inclusion of $C+1$ in the sample space of c_t supports a flexible, open-ended creation of new memories—a hallmark of nonparametric models.

If $c_t = C+1$ (i.e. the context is new), C is incremented and $\boldsymbol{\beta}$ is transformed by sampling $b \sim \mathrm{Beta}(1, \gamma)$ and assigning $\beta_{C+1} \leftarrow (1-b)\beta_C$ and $\beta_C \leftarrow b\beta_C$. Similarly, if $q_t = Q+1$ (i.e. the sensory cue is new), Q is incremented and $\boldsymbol{\beta}^{\mathrm{e}}$ is transformed by sampling $b^{\mathrm{e}} \sim \mathrm{Beta}(1, \gamma^{\mathrm{e}})$ and assigning $\beta^{\mathrm{e}}_{Q+1} \leftarrow (1-b^{\mathrm{e}})\beta^{\mathrm{e}}_Q$ and $\beta^{\mathrm{e}}_Q \leftarrow b^{\mathrm{e}}\beta^{\mathrm{e}}_Q$.

2.3.3 Propagate the sufficient statistics for the states

Conditioned on the sampled context variable, the sufficient statistics (mean and variance) for the state of each known context are propagated via the measurement update equations of the Kalman filter (Algorithm 3).

```
for $j = 1, \ldots, C$ do
  if $c_t = j$ then
    $e_t^{(j)} = y_t - \hat{y}_t^{(j)}$
    $k_t^{(j)} = v_{t|t-1}^{(j)} p_t^{(j)}$
    $\hat{x}_{t|t}^{(j)} = \hat{x}_{t|t-1}^{(j)} + k_t^{(j)} e_t^{(j)}$
    $v_{t|t}^{(j)} = (1 - k_t^{(j)}) v_{t|t-1}^{(j)}$
  else
    $\hat{x}_{t|t}^{(j)} = \hat{x}_{t|t-1}^{(j)}$
    $v_{t|t}^{(j)} = v_{t|t-1}^{(j)}$
  end if
end for
```

Algorithm 3: State filtering. The difference between the actual state feedback y_t and the predicted state feedback $\hat{y}_t^{(j)}$ (i.e. the prediction error $e_t^{(j)}$) for context j is used to update the mean of the predicted state distribution $\hat{x}_{t|t-1}^{(j)}$ of that context if it is inferred to be responsible for generating the state feedback (i.e. if $c_t = j$). The prediction error is scaled by the Kalman gain $k_t^{(j)}$, which is close to 0 when $\sigma_\mathrm{r}^2 \gg v_{t|t-1}^{(j)}$ and close to 1 when $\sigma_\mathrm{r}^2 \ll v_{t|t-1}^{(j)}$. Note that although a single particle updates the state of only one context in an all-or-none manner (the state of the context sampled by that particle), different particles may update the states of different contexts, thus leading to graded updates on average across the ensemble of particles.

2.3.4 Propagate the sufficient statistics for the state retention and drift parameters

For each $j \in \{1, \ldots, C\}$, a pair of states $(x_{t-1}^{(j)}, x_t^{(j)})$ are sampled from

$$\begin{aligned} p(x_{t-1}^{(j)}, x_t^{(j)} | c_t, z_{t-1}, y_t) &= p(x_{t-1}^{(j)} | c_t, z_{t-1}, y_t) p(x_t^{(j)} | x_{t-1}^{(j)}, c_t, z_{t-1}, y_t) \\ &= \mathcal{N}(\hat{x}_{t-1|t}^{(j)}, v_{t-1|t}^{(j)}) \mathcal{N}(\tilde{x}_{t|t}^{(j)}, \tilde{v}_{t|t}^{(j)}), \end{aligned} \tag{S19}$$

where $\hat{x}_{t-1|t}^{(j)} = \hat{x}_{t-1|t-1}^{(j)} + g(\hat{x}_{t|t}^{(j)} - \hat{x}_{t|t-1}^{(j)})$, $g = a^{(j)} v_{t-1|t-1}^{(j)} / v_{t|t-1}^{(j)}$, $v_{t-1|t}^{(j)} = v_{t-1|t-1}^{(j)} + g^2 (v_{t|t}^{(j)} - v_{t|t-1}^{(j)})$, $\tilde{x}_{t|t}^{(j)} = \tilde{v}_{t|t}^{(j)} [(a^{(j)} x_{t-1}^{(j)} + d^{(j)}) / \sigma_\mathrm{q}^2 + \delta_{c_t j} y_t / \sigma_\mathrm{r}^2]$ and $\tilde{v}_{t|t}^{(j)} = 1/(1/\sigma_\mathrm{q}^2 + \delta_{c_t j} / \sigma_\mathrm{r}^2)$. The sufficient statistics for the state retention and drift parameters are then propagated as follows:

$$\begin{aligned} s_1^{(j)} &\leftarrow s_1^{(j)} + x_t^{(j)} \bar{\boldsymbol{x}}_{t-1}^{(j)} \\ s_2^{(j)} &\leftarrow s_2^{(j)} + \bar{\boldsymbol{x}}_{t-1}^{(j)} \bar{\boldsymbol{x}}_{t-1}^{(j)\mathsf{T}}, \end{aligned} \tag{S20}$$

where $\bar{\boldsymbol{x}}_{t-1}^{(j)} = \left[x_{t-1}^{(j)} \; 1 \right]^\mathsf{T}$.

2.3.5 Propagate the sufficient statistics for the parameters governing the global transition and cue probabilities

The context transition counts and the context-specific cue counts are propagated by incrementing $n_{c_{t-1} c_t}$ and $n_{c_t q_t}^\mathrm{e}$, respectively.

2.3.6 Sample the state retention and drift parameters

For each $j \in \{1, \dots, C\}$, the hyperparameters of the posterior distribution of the state retention and drift parameters are computed:

$$\begin{aligned} \boldsymbol{\mu}^{(j)} &= \boldsymbol{\Sigma}^{(j)}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + s_1^{(j)}/\sigma_{\mathrm{q}}^2) \\ \boldsymbol{\Sigma}^{(j)} &= (\boldsymbol{\Sigma}^{-1} + s_2^{(j)}/\sigma_{\mathrm{q}}^2)^{-1}, \end{aligned} \tag{S21}$$

and a new set of state retention and drift parameters are sampled from

$$\boldsymbol{\omega}^{(j)} \mid \boldsymbol{\mu}^{(j)}, \boldsymbol{\Sigma}^{(j)} \ \sim \ \mathcal{TN}(\boldsymbol{\mu}^{(j)}, \boldsymbol{\Sigma}^{(j)}). \tag{S22}$$

2.3.7 Sample the global cue probabilities

To sample $\boldsymbol{\beta}^{\mathrm{e}}$, a Chinese restaurant process is first simulated to sample each m_{jk} (the number of tables in restaurant j serving dish k). For each $j \in \{1, \dots, C\}$ and $k \in \{1, \dots, Q\}$, m_{jk} and n are initialised to 0. Then, for $i = 1, \dots, n_{jk}^{\mathrm{e}}$ (i.e. for each customer in restaurant j eating dish k), a sample is drawn from

$$x \sim \mathrm{Bernoulli}\left(\frac{\alpha^{\mathrm{e}}\beta_k^{\mathrm{e}}}{n + \alpha^{\mathrm{e}}\beta_k^{\mathrm{e}}}\right), \tag{S23}$$

n is incremented, and if $x = 1$, m_{jk} is incremented.

Conditioned on each $m_k^{\mathrm{d}} = \sum_j m_{jk}$ (the total number of tables in all restaurants serving dish k), the global cue distribution is sampled from

$$(\beta_1^{\mathrm{e}}, \dots, \beta_Q^{\mathrm{e}}, \beta_{Q+1}^{\mathrm{e}}) \sim \mathrm{Dirichlet}(m_1^{\mathrm{d}}, \dots, m_Q^{\mathrm{d}}, \gamma^{\mathrm{e}}). \tag{S24}$$

2.3.8 Sample the global transition probabilities

To sample $\boldsymbol{\beta}$, a Chinese restaurant process is first simulated to sample each m_{jk} (the number of tables in restaurant j *serving* dish k). For each $(j, k) \in \{1, \dots, C\}^2$, m_{jk} and n are initialised to 0. Then, for $i = 1, \dots, n_{jk}$ (i.e. for each customer in restaurant j eating dish k), a sample is drawn from

$$x \sim \mathrm{Bernoulli}\left(\frac{\alpha\beta_k + \kappa\delta_{jk}}{n + \alpha\beta_k + \kappa\delta_{jk}}\right), \tag{S25}$$

n is incremented, and if $x = 1$, m_{jk} is incremented.

Then, for each $j \in \{1, \dots, C\}$, w_j (the number of times a dish considered at a new table in restaurant j is overridden by the specialty dish) is sampled from

$$w_j \sim \mathrm{Binomial}\left(m_{jj}, \frac{\rho}{\rho + \beta_j(1 - \rho)}\right). \tag{S26}$$

Finally, each $\bar{m}_{jk}$ (the number of tables in restaurant j *considering* dish k) is obtained as

$$\bar{m}_{jk} = \begin{cases} m_{jj} - w_j & \text{if } j = k \\ m_{jk} & \text{otherwise.} \end{cases} \tag{S27}$$

Conditioned on each $\bar{m}_k^{\mathrm{d}} = \sum_j \bar{m}_{jk}$ (the total number of tables in all restaurants considering dish k), the global transition distribution is sampled from

$$(\beta_1, \dots, \beta_C, \beta_{C+1}) \sim \mathrm{Dirichlet}(\bar{m}_1^{\mathrm{d}}, \dots, \bar{m}_C^{\mathrm{d}}, \gamma). \tag{S28}$$

2.4 Stationary context distribution

In Extended Data Fig. 1c and Extended Data Fig. 9c-d, we plot the probability of each context under the stationary distribution, which represents the expected frequency of each context in the long run. We calculate the stationary context distribution by solving $\psi = \psi\hat{\mathbf{\Pi}}$ for ψ (a row vector) subject to the constraint that ψ is a valid probability distribution (i.e. all elements of ψ are non-negative and sum to 1). Here $\hat{\mathbf{\Pi}}$ is the expected local transition probability matrix with elements given by

$$\hat{\pi}_{jk} = \frac{1}{P}\sum_{i=1}^{P}\frac{\alpha\beta_k^{(i)} + \kappa\delta_{jk} + n_{jk}^{(i)}}{\alpha + \kappa + n_{j\cdot}^{(i)}}, \tag{S29}$$

where i indexes each particle.

2.5 Validation of inference

We validated our approximate inference algorithm on synthetic data generated under the generative model. Data were generated in two settings that differed in terms of the upper bound on the number of possible contexts (determined by the truncation level of the stick-breaking process). In the single-context setting, only one context was possible. In the multiple-context setting, up to 10 contexts were possible. For the single-context and multiple-context settings, we generated 4000 and 2000 synthetic data sets, respectively, of 500 time steps each. The parameters and hyperparameters used to generate these data sets were chosen so that the distributions of the numbers of contexts and cues (Extended Data Fig. 2b) were typical for motor learning experiments.

Each data set consisted of a sequence of time-varying latent variables (contexts and states) and observations (state feedback and sensory cues) as well as a set of time-invariant parameters for the state dynamics of each context (state retention factor and state drift). We applied our inference algorithm with 100 particles to the sequence of observations and at each time step calculated a posterior predictive p-value for each of the time-varying latent variables, observations and parameters. For continuous variables (state feedback, states and parameters), the posterior predictive p-value was calculated by evaluating the cumulative distribution function (CDF) of the predictive distribution at the true value of the variable. For discrete variables with integer-valued support (contexts and sensory cues), the posterior predictive p-value was calculated as

$$\text{posterior predictive p-value} = F(x-1) + uf(x), \tag{S30}$$

where $f()$ is the predicted probability mass function, $F()$ is the cumulative mass function, x is the true value of the variable and $u \sim \mathcal{U}(0,1)$ is a uniform random variable on $[0,1]$. Crucially, if the predictive distributions/functions are well calibrated, the distributions of posterior predictive p-values will be uniformly distributed between 0 and 1, and hence the cumulative probability of posterior predictive p-values will lie on the identity line (Extended Data Fig. 2a-b).

The label-switching problem (see Section 2.9.2) complicates model validation with respect to variables that are associated with a specific context (the state of each context, the parameters for the state of each context and the context itself). We addressed this issue in several ways. In the single-context setting (Extended Data Fig. 2a), we circumvented the label-switching problem by limiting the number of contexts to 1. In the multiple-context setting (Extended Data Fig. 2b), we either integrated out the context or optimally permuted the context labels. Specifically, to calculate the posterior predictive p-value for the state, we evaluated the CDF of the marginal predictive distribution (the sum over contexts of the predictive distributions of the state of each context weighted by the predicted context probabilities) at the true value of the state of the current context. An analogous approach was taken to calculate the posterior predictive p-value for the parameters of the state. To calculate the posterior predictive p-value for the context, we first found the optimal permutation of labels that minimised the Hamming distance between the context sequence of each particle and the true context sequence. This relabelling procedure was repeated at each time step based on the sequence of contexts up to and including the current time step.

2.6 Extension to visuomotor rotation paradigms

In visuomotor rotation experiments, the cursor moves in a different direction to the hand (which is occluded from vision). This introduces a discrepancy between the location of the hand as perceived by vision and proprioception. To model this discrepancy, we include a context-dependent bias parameter $b^{(c_t)}$ in the state feedback (Eq. 5):

$$y_t = x_t^{(c_t)} + b^{(c_t)} + v_t, \qquad v_t \sim \mathcal{N}(0, \sigma_\mathrm{r}^2). \tag{S31}$$

To support Bayesian inference, we place a normal distribution prior over this parameter:

$$b^{(j)} \mid \mu_\mathrm{b}, \sigma_\mathrm{b} \sim \mathcal{N}(\mu_\mathrm{b}, \sigma_\mathrm{b}^2). \tag{S32}$$

We set μ_b to zero based on the assumption that positive and negative biases are equally probable and σ_b to 70^{-1} by hand to match the empirical data in Extended Data Fig. 9e-l.

The inference algorithm is extended in the following ways:

1. For each $j \in \{1, \ldots, C\}$, the sufficient statistics for the bias parameter are propagated as follows:

$$\begin{aligned} s_3^{(j)} &\leftarrow s_3^{(j)} + \delta_{c_t j}(y_t - x_t^{(j)}) \\ s_4^{(j)} &\leftarrow s_4^{(j)} + \delta_{c_t j}, \end{aligned} \tag{S33}$$

   where $x_t^{(j)}$ is sampled from $p(x_t^{(j)}|c_t, z_{t-1}, y_t) = \mathcal{N}(\hat{x}_{t|t}^{(j)}, v_{t|t}^{(j)})$. Note that this step is omitted on channel trials, as there is no state feedback.

2. For each $j \in \{1, \ldots, C\}$, the hyperparameters of the posterior distribution of the bias parameter are computed:

$$\begin{aligned} \mu_\mathrm{b}^{(j)} &= \sigma_\mathrm{b}^{(j)2}(\mu_\mathrm{b}/\sigma_\mathrm{b}^2 + s_3^{(j)}/\sigma_\mathrm{r}^2) \\ \sigma_\mathrm{b}^{(j)2} &= (1/\sigma_\mathrm{b}^2 + s_4^{(j)}/\sigma_\mathrm{r}^2)^{-1}. \end{aligned} \tag{S34}$$

   and a new bias parameter is sampled from:

$$b^{(j)} \mid \mu_\mathrm{b}^{(j)}, \sigma_\mathrm{b}^{(j)} \ \sim \ \mathcal{N}(\mu_\mathrm{b}^{(j)}, \sigma_\mathrm{b}^{(j)2}). \tag{S35}$$

The inference algorithm is also modified in the following ways:

1. The predicted state feedback (Algorithm 2) is changed to $\hat{y}_t^{(j)} = \hat{x}_{t|t-1}^{(j)} + b^{(j)}$.

2. The mean of the state distribution used to propagate the sufficient statistics for the state (Eq. S19) is changed to $\tilde{x}_{t|t}^{(j)} = \tilde{v}_{t|t}^{(j)}[(a^{(j)}x_{t-1}^{(j)} + d^{(j)})/\sigma_\mathrm{q}^2 + \delta_{c_t j}(y_t - b^{(j)})/\sigma_\mathrm{r}^2]$.

2.7 Model implementation

We applied the inference algorithm described in Section 2.3 to a sequence of noisy state feedback observations and, where applicable, sensory cues (numbered by the order they were presented in the experiment). On each trial, the state feedback was assigned a value of 0 (null-field trials), $+1$ (P$^+$ perturbation trials) or -1 (P$^-$ perturbation trials) plus i.i.d. zero-mean Gaussian observation noise with variance σ_r^2. Because both motor noise and sensory noise influence observed movement kinematics (the state feedback), we set σ_r^2 to $\sigma_\mathrm{m}^2 + \sigma_\mathrm{s}^2$ under the assumption that motor and sensory noise are i.i.d. Gaussian variables (with variances σ_m^2 and σ_s^2, respectively) that sum to produce the final observation noise. To reduce the number of free parameters in the model, we set σ_s to 0.03 under the assumption that sensory noise is typically no more than around one tenth ($\sim$ 3 s.d.) of the perturbation magnitude.

Adaptation a_t on trial t was modelled as the motor output u_t plus i.i.d. zero-mean Gaussian noise:

$$a_t \sim \mathcal{N}(u_t, \sigma_\mathrm{m}^2). \tag{S36}$$

The motor output was obtained by summing, for each particle, the means of the predicted state feedback distributions for each known context and a novel context weighted by their predicted probabilities and then averaging across all P particles:

$$u_t = \frac{1}{P}\sum_{i=1}^{P}\sum_{j=1}^{C+1}\left[\int y_t p(y_t|c_t = j, z_{t-1}^{(i)})\mathrm{d}y_t\right] p(c_t = j|q_t, z_{t-1}^{(i)}), \tag{S37}$$

where the predicted state feedback distribution $p(y_t|c_t, z_{t-1})$ is given in Eq. S17, and the predicted probabilities $p(c_t|q_t, z_{t-1})$ are

$$\begin{aligned} p(c_t|q_t, z_{t-1}) &\propto p(q_t, c_t|z_{t-1}) \\ &= p(q_t|c_t, z_{t-1})p(c_t|z_{t-1}), \end{aligned} \tag{S38}$$

where $p(q_t|c_t, z_{t-1})$ and $p(c_t|z_{t-1})$ are defined in Eqs. S15 and S16. Note that in the absence of a bias parameter (or when the bias parameter is zero), the mean of the predicted state feedback distribution for each context is equal to the mean of the predicted state distribution for each context. Hence, Eq. S37 is applicable to force field experiments, where there is no bias and the motor output can be defined in terms of the predicted state distribution for each context, as well as visuomotor rotation experiments, where there is a bias and the motor output is defined in terms of the predicted state feedback distribution for each context.

On channel trials (P$^\mathrm{c}$), we omitted state feedback as the state (e.g. the magnitude of a force field or visuomotor rotation) is not observed. This was achieved by modifying the inference algorithm in the following ways:

1. The state feedback likelihood term (Eq. S17) did not contribute to the weights used to resample particles (Eq. S14) or the probabilities used to propagate the context (Eq. S18). If there were no sensory cues, resampling was omitted altogether as the particle weights are uniform in this case.

2. The measurement update steps were omitted when updating the state estimate (Algorithm 3), that is $\hat{x}_{t|t}^{(j)} = \hat{x}_{t|t-1}^{(j)}$ and $v_{t|t}^{(j)} = v_{t|t-1}^{(j)}$. Hence, on channel trials, each state estimate is updated based only on the inferred dynamics (state retention and drift parameters) ascribed to that context, and there is no error-based learning.

3. To propagate the sufficient statistics for the retention and drift parameters, a pair of states $(x_{t-1}^{(j)}, x_t^{(j)})$ are sampled from a distribution that is equivalent to Eq. S19 but that does not condition on y_t (and

hence does not need to condition on c_t either):

$$
\begin{aligned}
p(x_{t-1}^{(j)}, x_t^{(j)} | z_{t-1}) &= p(x_{t-1}^{(j)} | z_{t-1}) p(x_t^{(j)} | x_{t-1}^{(j)}, z_{t-1}) \\
&= \mathcal{N}(\hat{x}_{t-1|t-1}^{(j)}, v_{t-1|t-1}^{(j)}) \mathcal{N}(\tilde{x}_{t|t-1}^{(j)}, \tilde{v}_{t|t-1}^{(j)}),
\end{aligned}
\tag{S39}
$$

where $\tilde{x}_{t|t-1}^{(j)} = a^{(j)} x_{t-1}^{(j)} + d^{(j)}$ and $\tilde{v}_{t|t-1}^{(j)} = \sigma_{\mathrm{q}}^2$.

The number of possible contexts in the COIN model—although infinite in principle—was limited to be finite in practice. This was achieved by truncating the stick-breaking process of the GEM to a finite level. In most instances, we limited the number of possible contexts to 10, as this number was greater than the true number of contexts in the experiments we modelled (typically 2-3). In the single-context setting of model validation, we limited the number of possible contexts to 1 (see Validation of the COIN model). Note that when the number of possible contexts is limited to 1, the nonparametric switching state-space model reduces to a single context (i.e. non-switching) state-space model. Moreover, if the parameters of the state dynamics are also known (i.e. not learned online), the Kalman filter is recovered as a special case of the inference algorithm of the COIN model.

To reduce the number of free parameters in the model, we set $\gamma = \gamma^{\mathrm{e}} = 0.1$, except during model validation, where we set $\gamma = \gamma^{\mathrm{e}} = 0.3$ to generate distributions of contexts and cues that are typical of motor learning experiments.

The algorithm was initialised with $C = 0, Q = 0, \beta_1 = 1, \beta_1^{\mathrm{e}} = 1$ and the sufficient statistics for the parameters set to 0.

2.8 Model fitting

2.8.1 Objectives and optimiser

In both experiments, we fit the COIN model to the data of individual participants by fitting the set of parameters ϑ so as to maximise the data log likelihood. In the spontaneous/evoked recovery experiment, $\vartheta = \{\sigma_{\mathrm{q}}, \mu_{\mathrm{a}}, \sigma_{\mathrm{a}}, \sigma_{\mathrm{d}}, \alpha, \rho, \sigma_{\mathrm{m}}\}$, and in the memory updating experiment, which included sensory cues, an additional parameter was fit so that $\vartheta = \{\sigma_{\mathrm{q}}, \mu_{\mathrm{a}}, \sigma_{\mathrm{a}}, \sigma_{\mathrm{d}}, \alpha, \rho, \sigma_{\mathrm{m}}, \alpha^{\mathrm{e}}\}$. The likelihood is

$$
p(\{a_{t'}\}_{t' \in \mathcal{T}} | x_{1:T}^{\star}, q_{1:T}, \vartheta) = \iint p(\{a_{t'}\}_{t' \in \mathcal{T}} | \mathcal{Z}_{1:T}, \vartheta) p(\mathcal{Z}_{1:T} | y_{1:T}, q_{1:T}, \vartheta) p(y_{1:T} | x_{1:T}^{\star}, \vartheta) \, \mathrm{d}\mathcal{Z}_{1:T} \, \mathrm{d}y_{1:T}. \tag{S40}
$$

Here $\{a_{t'}\}_{t' \in \mathcal{T}}$ is the adaptation data (noisy motor output) of the participant on the set of trials $\mathcal{T}$ that were channel trials, $\mathcal{Z}_{1:T}$ (where $\mathcal{Z}_T = \{z_T^{(i)}\}_{i=1}^P$, see Section 2.3) is the sequence of inferences made by the COIN model, $x_{1:T}^{\star}$ and $q_{1:T}$ are the experimental perturbations and sensory cues (if applicable) presented to the participant, respectively, and $y_{1:T}$ is the sequence of state feedback observations (perturbations plus observation noise) experienced by the participant. Note that the state feedback is integrated out, as the actual observation noise that the participant experienced is hidden from the perspective of the experimenter. We approximate the likelihood using Monte Carlo integration:

$$
\begin{aligned}
p(\{a_{t'}\}_{t' \in \mathcal{T}} | x_{1:T}^{\star}, q_{1:T}, \vartheta) &\approx \frac{1}{R} \sum_{r=1}^{R} p(\{a_{t'}\}_{t' \in \mathcal{T}} | \mathcal{Z}_{1:T}^{(r)}, \vartheta) \qquad \mathcal{Z}_{1:T}^{(r)} \sim p(\mathcal{Z}_{1:T} | x_{1:T}^{\star}, q_{1:T}, \vartheta) \\
&= \frac{1}{R} \sum_{r=1}^{R} \prod_{t' \in \mathcal{T}} p(a_{t'} | \mathcal{Z}_{t'}^{(r)}, \vartheta),
\end{aligned}
\tag{S41}
$$

where each $\mathcal{Z}_{1:T}^{(r)} \sim p(\mathcal{Z}_{1:T}|x_{1:T}^{\star}, q_{1:T}, \vartheta)$ is obtained by running the COIN model conditioned on a state feedback sequence $y_{1:T}^{(r)}$ sampled from $p(y_{1:T}|x_{1:T}^{\star}, \vartheta) = \prod_{t=1}^{T} \mathcal{N}(x_t^{\star}, \sigma_{\mathrm{r}}^2)$ and $p(a_{t'}|\mathcal{Z}_{t'}^{(r)}, \vartheta) = \mathcal{N}(u_{t'}^{(r)}, \sigma_{\mathrm{m}}^2)$. Note that this objective is stochastic because we sample observation noise to generate the state feedback. Consequently, to fit the COIN model, we used Bayesian adaptive direct search (BADS)[13], a Bayesian optimisation algorithm that alternates between a series of fast, local Bayesian optimisation steps and a systematic, slower exploration of a mesh grid. Optimisation was performed from 30 random initial parameter settings with $P = 100$ particles and $R = 100$ 'runs'. Once each optimisation was complete, we re-calculated the log likelihood using $P = 1000$ particles and $R = 1000$ runs to obtain a lower-variance estimate of the log likelihood. This estimate was used to choose the best fit out of 30 for each participant.

To fit the COIN model to group average data (spontaneous/evoked recovery experiment), we defined the likelihood as

$$p(\{\bar{a}_i\}_{i=1}^{N}|x_{1:T}^{\star}, \vartheta) \approx \frac{1}{R}\sum_{r=1}^{R}\prod_{i=1}^{N}\mathcal{N}(\bar{a}_i|\bar{u}_i^{(r)}, \sigma_{\mathrm{m}}^2/S), \tag{S42}$$

where $\bar{a}_i$ and $\bar{u}_i$ are the average adaptation data and motor output of the COIN model across participants on channel trial i (channel trials are numbered according to the order of their presentation), respectively, and S is the number of participants in the group. The motor output of the COIN model for each participant was obtained by running the COIN model conditioned on a participant-specific state feedback sequence sampled from $p(y_{1:T}|x_{1:T}^{\star}, \vartheta)$. Note that the variance of the motor noise is scaled by $1/S$, as the variance of the mean of S independent random variables each with variance σ_{m}^2 is σ_{m}^2/S. To fit the model to the average spontaneous recovery group data and the average evoked recovery group data using the same set of parameters, we calculated the log likelihood separately for each group (using Eq. S42) and optimised the sum of the log likelihoods.

2.8.2 Validation of fitting: parameter recovery

We used the parameters from the fits of the COIN model to the data for each participant in the spontaneous recovery and evoked recovery experiments to generate 10 synthetic data sets for each participant from the corresponding experiment. For a given set of parameters in a given experiment, there were two sources of variability across different synthetic data sets: sensory noise and motor noise. We then fit each synthetic data set with the COIN model as we did with real data.

For parameter recovery (Extended Data Fig. 2c), we compared the COIN model parameters that were used to generate the synthetic data ('true' parameters) with the COIN model parameters fit to these synthetic data sets.

2.9 Inferring internal representations

2.9.1 Integrating out observation noise

The actual observation noise (and thus state feedback) that a participant perceives is hidden from the perspective of the experimenter. Therefore, to infer the internal representations of a participant (their sequence of inferences), rather than perform one 'run' of a simulation, conditioned on one particular state feedback sequence, we perform $R = 100$ runs, each conditioned on a different state feedback sequence (sampled from the prior). We then integrate out the state feedback by computing a weighted average of the runs, with each run assigned a weight based on the likelihood of the adaptation data of the participant (if available). Formally, this corresponds to approximating the expected value of a function f of the state feedback (e.g. the internal representations of the COIN model) using importance sampling:

$$\mathbb{E}_{p(y_{1:T}|\{a_{t'}\}_{t'\in\mathcal{T}},x^{\star}_{1:T},q_{1:T},\hat{\vartheta}))}[f(y_{1:T})] \approx \sum_{r=1}^{R} \frac{w_T^{(r)}}{\sum_{r'=1}^{R} w_T^{(r')}} f(y_{1:T}^{(r)}) \qquad y_{1:T}^{(r)} \sim p(y_{1:T}|x^{\star}_{1:T},\hat{\vartheta}), \tag{S43}$$

where $p(y_{1:T}|x^{\star}_{1:T},\hat{\vartheta})$ is the prior distribution of the state feedback, $p(y_{1:T}|\{a_{t'}\}_{t'\in\mathcal{T}},x^{\star}_{1:T},q_{1:T},\hat{\vartheta})$ is the posterior distribution of the state feedback, $\{a_{t'}\}_{t'\in\mathcal{T}}$ is the adaptation data of the participant on the set of trials $\mathcal{T}$ that were channel trials, $x^{\star}_{1:T}$ and $q_{1:T}$ are the sequences of perturbations and sensory cues (where applicable) presented to the participant and $\hat{\vartheta}$ are the COIN model parameters fit to the data of the participant (here and elsewhere in this section we use the hat notation to indicate a maximum likelihood estimate). The importance weights are equal to the joint likelihood of the adaptation data measured so far:

$$w_T^{(r)} = p(\{a_{t'}\}_{t'\in\mathcal{T}}|y_{1:T}^{(r)},q_{1:T},\hat{\vartheta}). \tag{S44}$$

Thus runs that place higher probability on the adaptation data are assigned greater importance weights. If we assume that the sequence of inferences made by the COIN model $\mathcal{Z}_{1:T}^{(r)}$ (where $\mathcal{Z}_T^{(r)} = \{z_T^{(i,r)}\}_{i=1}^{P}$, see Section 2.3) is a deterministic function of the state feedback and sensory cue observations, which is true if the number of particles in the model is sufficiently large, the importance weights become

$$w_T^{(r)} = p(\{a_{t'}\}_{t'\in\mathcal{T}}|\mathcal{Z}_{1:T}^{(r)},\hat{\vartheta}) \qquad \mathcal{Z}_{1:T}^{(r)} \sim p(\mathcal{Z}_{1:T}|y_{1:T}^{(r)},q_{1:T},\vartheta). \tag{S45}$$

These weights can be obtained by running the COIN model (conditioned on the state feedback and sensory cue observations) to generate the sequence of inferences. Note that these inferences are the quantities we are actually interested in and which we plot (see below)—the state feedback, in contrast, is a nuisance variable. Importantly, the importance weights are computed as a product of densities, $p(\{a_{t'}\}_{t'\in\mathcal{T}}|\mathcal{Z}_{1:T}^{(r)},\hat{\vartheta}) = \prod_{t'\in\mathcal{T}} p(a_{t'}|\mathcal{Z}_{t'}^{(r)},\hat{\vartheta})$, which for a large number of factors can result in only a few runs having significant weight, a problem known as degeneracy. To identify a set of weights that is close to being degenerate, we calculate the effective sample size:

$$n_T^{\text{eff}} = \frac{\left(\sum_{r=1}^{R} w_T^{(r)}\right)^2}{\sum_{r=1}^{R} w_T^{(r)2}}. \tag{S46}$$

The effective sample size is equal to R when all runs have equal weight and 1 when only one run has nonzero weight. To avoid degeneracy, we calculate the importance weights sequentially for $t = 1,\ldots,T$ and resample runs with probabilities proportional to their weights whenever the effective sample size falls below a threshold of $R/2$. Resampling resets the weights to $1/R$, thus avoiding degeneracy.

Here we describe the sequential importance sampling with resampling (particle filtering) algorithm for a single trial. At the end of trial $t-1$, each run is associated with a state feedback sequence $y_{1:t-1}^{(r)}$, a sequence of inferences $\mathcal{Z}_{1:t-1}^{(r)}$ obtained by running the COIN model conditioned on the state feedback sequence and a weight $w_{t-1}^{(r)}$. On trial t, the state feedback $y_t^{(r)}$ is sampled from $p(y_t|x_t^{\star},\hat{\vartheta})$, and $\mathcal{Z}_{t-1}^{(r)}$ is propagated by sampling $\mathcal{Z}_t^{(r)}$ from $p(\mathcal{Z}_t|\mathcal{Z}_{t-1}^{(r)},y_t^{(r)},q_t,\hat{\vartheta})$; a sample of $\mathcal{Z}_t^{(r)} \sim p(\mathcal{Z}_t|\mathcal{Z}_{t-1}^{(r)},y_t^{(r)},q_t,\hat{\vartheta})$ is obtained by performing one update of the COIN model from trial $t-1$ to t conditioned on the state feedback sample. The trajectories $y_{1:t-1}^{(r)}$ and $\mathcal{Z}_{1:t-1}^{(r)}$ are then augmented with $y_t^{(r)}$ and $\mathcal{Z}_t^{(r)}$, respectively. Next the importance weights are updated. If an adaptation measurement was made on trial t (e.g. the trial was a channel trial), the weights are updated according to $w_t^{(r)} = p(a_t|\mathcal{Z}_t^{(r)},\hat{\vartheta})w_{t-1}^{(r)}$, otherwise they are left unchanged as $w_t^{(r)} = w_{t-1}^{(r)}$. Finally, n_t^{eff} is calculated, and if $n_t^{\text{eff}} < 1/R$, the runs are resampled with probabilities proportional to their weights, and the weights are reset to $w_t^{(r)} = 1/R$.

```
propagate
  for $r = 1, \ldots, R$ do
    sample $y_t^{(r)} \sim p(y_t|x_t^\star, \hat{\vartheta})$, where $p(y_t|x_t^\star, \hat{\vartheta}) = \mathcal{N}(x_t^\star, \hat{\sigma}_\mathrm{r}^2)$, and then sample
    $\mathcal{Z}_t^{(r)} \sim p(\mathcal{Z}_t|\mathcal{Z}_{t-1}^{(r)}, y_t^{(r)}, q_t, \hat{\vartheta})$ and augment trajectories $y_{1:t-1}^{(r)}$ and $\mathcal{Z}_{1:t-1}^{(r)}$ with $y_t^{(r)}$ and $\mathcal{Z}_t^{(r)}$
  end for
weight
  for $r = 1, \ldots, R$ do
    if an adaptation measurement was made on trial $t$ then
      $w_t^{(r)} = p(a_t|\mathcal{Z}_t^{(r)}, \hat{\vartheta}) w_{t-1}^{(r)}$, where $p(a_t|\mathcal{Z}_t^{(r)}, \hat{\vartheta}) = \mathcal{N}(u_t^{(r)}, \hat{\sigma}_\mathrm{m}^2)$
    else
      $w_t^{(r)} = w_{t-1}^{(r)}$
    end if
  end for
resample
  if $n_t^\mathrm{eff} < R/2$ then
    resample runs with probabilities proportional to $w_t^{(r)}$ and reset weights to $1/R$
  end if
```

Algorithm 4: The particle filtering algorithm for a single trial, trial t. State feedback is sampled from the prior and then weighted by the likelihood. The weighted samples of state feedback allow the state feedback to be integrated out of functions that depend on it (Eq. S43), such as inferences made by the COIN model.

The particle filtering algorithm is summarised in Algorithm 4 for a single trial. Note that this algorithm involves a two-level hierarchy of particle methods, as each particle in the ensemble $\{\mathcal{Z}_t^{(r)}\}_{r=1}^R$ is itself an ensemble of particles, that is $\mathcal{Z}_t^{(r)} = \{z_t^{(i,r)}\}_{i=1}^P$. At the bottom level of the hierarchy, particle learning is used to simulate inference from the perspective of the participant conditioned on the state feedback and sensory cues (the propagate step of Algorithm 4 for one run is equivalent to Algorithm 1), while at the top level of the hierarchy, particle filtering is used to simulate inference from the perspective of the experimenter conditioned on the participant's adaptation data (Algorithm 4 in its entirety).

The end result of the particle filtering algorithm is a set of R runs, each associated with a state feedback sequence $y_{1:T}^{(r)}$, a sequence of inferences $\mathcal{Z}_{1:T}^{(r)}$ conditioned on the state feedback sequence and a weight $w_T^{(r)}$. In Fig. 2c,e, Fig. 3d, Extended Data Fig. 6f and Extended Data Fig. 7a-d, we plot the weighted average of the inferences associated with these runs, as per Eq. S43. Note that when no adaptation data is available, this algorithm reduces to performing R independent, equally-weighted simulations, each conditioned on a different state feedback sequence sampled from the prior. Such 'open-loop' simulations were used to generate the model data plotted in Fig. 1, Fig. 2b,d, Fig. 4, Extended Data Fig. 1c-e and Extended Data Figs. 4, 5, 8 and 9, which show the average inferences across runs (equally weighted).

2.9.2 Label-switching problem

Particle methods present a challenge when it comes to COIN model analysis, as contexts that have the same label (assigned based on the order in which they were sampled) in different particles do not necessarily correspond to the same ground-truth context. For example, in an experiment with two contexts, P$^+$ and P$^-$, one particle may assign most P$^+$ and P$^-$ trials to contexts 1 and 2, respectively, whereas another particle may assign most P$^+$ and P$^-$ trials to contexts 2 and 1, respectively. This so-called 'label-switching problem'[14], which arises because the likelihood is invariant under permutations of the context labels, renders context labels arbitrary.

We addressed the label-switching problem in two ways. In some instances (Fig. 4 and Extended Data Fig. 7c-d and Extended Data Fig. 8), we restricted our analysis to a single context (c^*) with the largest predicted probability or responsibility and thus disregarded the context labels. In other instances (Figs. 1 and 2, Extended Data Fig. 1c-e, Extended Data Fig. 2b for context variable only and Extended Data Figs. 5 and 9), we found the optimal permutation of labels that minimised the Hamming distance (number of label mismatches) between the context sequence of each particle in each run and the typical context sequence across all particles and runs. This was done on each trial based on the sequence of contexts sampled up to and including the current trial. The typical sequence was defined as the sequence with the minimum average Hamming distance to all other sequences. To calculate the Hamming distance between any two sequences, we first found the optimal (minimum Hamming distance) permutation of labels. For simplicity, we restricted this analysis to particles that had the same number of contexts as the most common number of contexts across particles and runs (i.e. the posterior mode).

Note that for variables that integrate out the context (adaptation, single-trial learning, the predicted state distribution, the predicted state feedback distribution, the inferred bias distribution), the label-switching problem does not exist. Hence, all particles were used to compute these variables.

2.10 Simulating existing data sets

We performed COIN model simulations on a diverse set of extant data in Fig. 4 (similarly Extended Data Figs. 5, 8 and 9) in a purely cross-validated manner, such that we used model parameters fitted to participants in our own experiments to make predictions for experiments conducted in other laboratories using other paradigms.

The paradigms in Fig. 4 and Extended Data Fig. 8 were simulated using the 40 sets of parameters fit to our individual participants' data from both experiments. One hundred simulations (each conditioned on a different noisy state feedback sequence) were performed for each parameter set. The results shown are based on the average of all of these simulations.

The paradigms in Extended Data Fig. 5a-o and Extended Data Fig. 9 were variations of the standard spontaneous recovery paradigm. Therefore, we simulated these paradigms (as well as the paradigm in Extended Data Fig. 5p-s) using the parameters fit to the average spontaneous and evoked recovery data sets. One hundred simulations (each conditioned on a different noisy state feedback sequence) were performed. The results shown are based on the average of these simulations.

2.10.1 Savings paradigm

We used the paradigm described in Ref. 20 to simulate savings in the COIN model (Fig. 4a). Participants completed two force-field learning sessions that were separated by a 5-minute break. Each session consisted of a pre-exposure phase (60 trials) with a null field (P^0). This was followed by an exposure phase (125 trials) with a velocity-dependent curl field (P^+). After the exposure phase, participants performed a counter-exposure phase (15 trials) with the opposite curl field (P^-). This was followed by a series of 50 channel trials (P^c). In addition, channel trials were randomly interspersed throughout the exposure phase (approximately 1 in every 10 trials).

2.10.2 Anterograde interference paradigm

We used the paradigm described in Ref. 3 to simulate anterograde interference in the COIN model (Fig. 4b). The paradigm consisted of a pre-exposure phase (160 trials) with a null field (P^0). This was followed by an exposure phase of variable length (13, 41, 112, 230, or 369 trials) with a velocity-dependent curl field (P^+). After the exposure phase, participants performed a counter-exposure phase (115 trials) with the opposite curl field (P^-). Channel trials were randomly interspersed throughout the exposure and counter-exposure phases (approximately 1 in every 7 trials).

2.10.3 Environmental consistency paradigms

We used the paradigm described in Experiment 1 in Ref. 2 to simulate the effect of environmental consistency on single-trial learning in the COIN model (Fig. 4c). The paradigm consisted of a pre-training phase (156 trials) with a null field (P^0) interspersed with triplets (1 in every 13 trials). This was followed by a training phase composed of 25 blocks (45 trials each). Each block of the training phase consisted of a sequence of 30 perturbation trials, 2 channel trials, 10 washout trials, and 1 triplet. During the perturbation trials, either P^+ or P^- was presented. Across trials, the perturbation either switched with probability $p(\text{switch})$ or remained the same with probability $p(\text{stay}) = 1 - p(\text{switch})$. Three groups performed the experiment with $p(\text{stay})$ set to 0.9, 0.5 and 0.1 for the groups who experienced the slowly, medium and rapidly switching environment, respectively.

As an additional demonstration of the effect of environmental consistency on single-trial learning in the COIN model (Extended Data Fig. 8), we simulated the P1N1, P1, P7, and P20 environments of the force-field adaptation task described in Ref. 21. The paradigm consisted of a pre-exposure phase (200 trials) with a null field (P^0). In the anti-consistent environment (P1N1), participants experienced 50 cycles each with a single P^+ trial, followed by a single P^- trial, followed by 11–13 P^0 trials. In the inconsistent environment (P1), participants experienced 45 cycles with a single P^+ trial, followed by 10-12 P^0 trials. In the moderately consistent environment (P7), participants experienced 27 cycles with seven P^+ trials, followed by 15–18 P^0 trials. In the highly consistent environment (P20), participants experienced 27 cycles with 20 P^+ trials, followed by 28–32 P^0 trials. To assess single-trial learning during exposure to the environments, channel trials were randomly interspersed before and after the first P^+ trial in a subset of the force-field cycles.

2.10.4 Extended exposure phase in a spontaneous recovery paradigm

We modified the spontaneous recovery paradigm (control condition of Experiment 2) described in Ref. 5 (see Section 2.10.8) to simulate the effect of extending the exposure phase on the amount of spontaneous recovery (Extended Data Fig. 5a-j). Following the manipulation of Ref. 13, we tripled the length of the exposure phase. Thus the number of exposure phase trials was increased from 200 trials (standard paradigm) to 600 trials (overlearning paradigm).

2.10.5 Pre-training in a spontaneous recovery paradigm

We used the paradigm described in Ref. 14 to simulate the effect of a pre-training phase on the amount of spontaneous recovery. There were three groups in the experiment, a standard group, a $\mathrm{P}^-_{\text{abrupt}}$ group and a $\mathrm{P}^-_{\text{gradual}}$ group (named Ab, BAb and $\mathrm{B_g}$Ab, respectively, in the authors' nomenclature). In the standard group, the paradigm consisted a pre-exposure phase (192 trials) with a null field (P^0). This was followed by an exposure phase (384 trials) with a velocity-dependent curl field (P^{+}). After the exposure phase, participants performed a counter-exposure phase (20 trials) with the opposite curl field (P^{-}). This was followed by a series of 364 channel trials (P^{c}). The paradigm in the $\mathrm{P}^{-}_{\text{abrupt}}$ and $\mathrm{P}^{-}_{\text{gradual}}$ groups was the same as in the standard group but with the following exceptions. In the $\mathrm{P}^{-}_{\text{abrupt}}$ group, a pre-training phase (384 trials) with the same curl field as in the counter-exposure phase (P^{-}) was inserted in between the pre-exposure and exposure phases. Similarly, in the $\mathrm{P}^{-}_{\text{gradual}}$ group, a pre-training phase (384 trials) with the same curl field as in the counter-exposure phase (P^{-}) was inserted in between the pre-exposure and exposure phases; however, in this group, the curl field was introduced gradually over 96 trials, maintained at full strength for 192 trials, and then gradually removed over 96 trials.

2.10.6 Abrupt vs. gradual introduction of a perturbation

The difference in deadaptation after abrupt vs. gradual introduction of a visuomotor rotation was studied in Ref. 17. The study examined intermanual transfer as well but here we simulate a similar paradigm without considering the transfer.

We simulated a paradigm that consisted of a pre-exposure phase (136 trials) in which cursor feedback was veridical (P^{0}). This was followed by an exposure phase (240 trials) in which the cursor was rotated either abruptly by 22.5° (P^{+}) or gradually in increments of $\sim 1^\circ$ every 10 trials. Finally, in a post-exposure phase (96 trials), the cursor feedback was again veridical (P^{0}).

2.10.7 Working memory and evoked recovery

We investigated the effect of a working memory task on contextual inference in the COIN model (Extended Data Fig. 9a-d) by simulating a force-field adaptation task (experiment 1) in Ref. 22. The paradigm consisted of a pre-exposure phase (192 trials) with a null field (P^{0}). This was followed by an exposure phase (384 trials) with a velocity-dependent curl field (P^{+}). After the exposure phase, participants performed a counter-exposure phase (20 trials) with the opposite curl field (P^{-}). Participants then completed either a memory task (memory group) or a non-memory task (non-memory group). This was followed by a series of 192 channel trials (P^{c}). Channel trials were randomly interspersed throughout the pre-exposure and exposure phases (1 in every 8 trials).

In the memory task, participants were shown 12 word pairs (e.g. “COMFORT-ATOM”, “LEGEND-BLANK”). Immediately after viewing the words, participants were then shown one word from each pair and instructed to say the corresponding word aloud. In the non-memory task, participants were shown strings of letters (e.g. “kdinedlr”) and were instructed to say aloud the number of vowels in each string.

We hypothesised that context probabilities, which are updated recursively in the COIN model, are maintained and updated in working memory. The effect of the working memory task is to erase these estimated probabilities from working memory so that participants instead infer the context based on the stationary distribution, which represents the expected frequency of each context. Therefore, on the first trial of the channel-trial phase (i.e. directly after the working memory task), we set the predicted probabilities to their values under the stationary distribution (calculated using the expected value of the transition probability matrix under the Dirichlet posterior). For the non-memory task, the COIN model was simulated as for a standard spontaneous recovery paradigm.

2.10.8 Explicit and implicit visuomotor learning

We investigated explicit and implicit learning in the COIN model (Extended Data Fig. 9e-l) by simulating a visuomotor rotation task (report and control condition of experiment 2) described in Ref. 5. The paradigm consisted of a pre-exposure phase (100 trials) in which cursor feedback was veridical (P^0). This was followed by an exposure phase (200 trials) in which the cursor was rotated by $45°$ in the clockwise direction (P$^+$, note we use this to represent a positive rotation in this visuomotor paradigm). After the exposure phase, participants performed a counter-exposure phase (20 trials) in which the cursor was rotated by $45°$ in the counterclockwise direction (P$^-$). This was followed by a series of 100 visual error clamp trials (P^c) in which the cursor moved straight to the target regardless of the participant's hand trajectory. During the pre-exposure, exposure and counter-exposure phases, the target was flanked by a $360°$ ring of numbered visual landmarks spaced $5.625°$ apart. Starting at trial 91 of the pre-exposure phase, participants were instructed to report verbally before each reach the landmark that they planned to push the manipulandum toward to make the cursor hit the target. These reported aiming directions were interpreted as the explicit component of learning. Implicit learning was quantified by subtracting the explicit component from the actual movement direction on each trial. After the end of the counter-exposure phase, participants were told to stop using any aiming strategy that they had developed and reach directly for the target during the remaining visual error clamp phase. A control group performed the identical paradigm but without any reporting of aim direction.

To simulate learning in a visuomotor rotation experiment in the COIN model, we included an additional parameter to reflect measurement bias (the difference between hand location perceived by proprioception and vision), which was inferred online (see Section 2.6).

3 Deterministic state-space models

3.1 Model definition

We also fit a class of n-rate deterministic state-space models to the data in the spontaneous/evoked recovery experiment. These models frame motor adaptation as sequential estimation of a task perturbation (e.g. the magnitude of a force field) using n separate adaptive states, each of which has its own own retention factor and learning rate. For the two-state (dual-rate) model, $n = 2$, and for the three-state model, $n = 3$. The individual states can be arranged into a state vector:

$$\hat{\boldsymbol{x}}_t = \left[\hat{x}_t^{(1)} \; \dots \; \hat{x}_t^{(n)}\right]^{\mathsf{T}}. \tag{S47}$$

The motor output on trial t is the sum of the elements in the state vector:

$$u_t = \sum_{i=1}^{n} \hat{x}_t^{(i)}. \tag{S48}$$

The error on trial t is the difference between the 'true' task perturbation $x_t^\star$ and the motor output:

$$e_t = x_t^\star - u_t. \tag{S49}$$

The task perturbation was 0 for null-field trials, $+1$ for P$^+$ field trials and -1 for P$^-$ field trials. The state vector is updated across trials as follows:

$$\hat{\boldsymbol{x}}_{t+1} = \boldsymbol{a} \odot \hat{\boldsymbol{x}}_t + \boldsymbol{b} e_t, \tag{S50}$$

where $\boldsymbol{a} = \begin{bmatrix} a^{(1)} & \ldots & a^{(n)} \end{bmatrix}^{\mathsf{T}}$ is a retention vector that governs trial-by-trial decay, $\boldsymbol{b} = \begin{bmatrix} b^{(1)} & \ldots & b^{(n)} \end{bmatrix}^{\mathsf{T}}$ is a learning-rate vector that governs error-dependent adaptation and $\odot$ denotes element-wise multiplication. For a task perturbation of +1, Eq. S50 can be rewritten as

$$\begin{aligned} \hat{\boldsymbol{x}}_{t+1} &= \boldsymbol{A}\hat{\boldsymbol{x}}_t + \boldsymbol{b} \\ &= (\boldsymbol{I}\boldsymbol{a} - \boldsymbol{B})\hat{\boldsymbol{x}}_t + \boldsymbol{b}, \end{aligned} \tag{S51}$$

where $\boldsymbol{I} \in \mathbb{R}^{n \times n}$ is the identity matrix and each column of $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ is equal to $\boldsymbol{b}$. We used this reparameterisation to ensure that fitted parameters lead to stable learning as assessed through the eigenvalues of $\boldsymbol{A}$.

3.2 Model fitting

In both experiments, we fit the deterministic state-space models to the data of individual participants by fitting the set of parameters ϑ so as to minimise the mean squared error between the model data (Eq. S48) and the adaptation data measured on channel trials. Under the assumption that the adaptation data is the model data plus i.i.d. Gaussian noise, $p(a_t|u_t, \sigma^2_{\mathrm{m}}) = \mathcal{N}(u_t, \sigma^2_{\mathrm{m}})$, this is equivalent to maximising the likelihood $p(\{a_{t'}\}_{t' \in \mathcal{T}}|x^{\star}_{1:T}, \vartheta) = \prod_{t' \in \mathcal{T}} p(a_{t'}|u_{t'}, \hat{\sigma}^2_{\mathrm{m}})$, where $\{a_{t'}\}_{t' \in \mathcal{T}}$ is the adaptation data (noisy motor output) of the participant on the set of trials $\mathcal{T}$ that were channel trials and $\hat{\sigma}^2_{\mathrm{m}}$ is the maximum likelihood estimate of the variance of the motor noise (the mean squared error). To ensure stable solutions, we constrained the eigenvalues of the matrix $\boldsymbol{A}$ in Eq. S51 to be between 0 and 1. Optimisation was performed from 30 random initial parameter settings using both MATLAB's fmincon and BADS. We report the best solution found by either optimiser.

4 Model comparison

4.1 Criterion

To perform model comparison for individual participants, we calculated the Bayesian information criterion[15]:

$$\text{BIC} = -2\log p(\{a_{t'}\}_{t' \in \mathcal{T}}|x^{\star}_{1:T}, \hat{\vartheta}) + k\log(N), \tag{S52}$$

where $\{a_{t'}\}_{t' \in \mathcal{T}}$ is the adaptation data (noisy motor output) of the participant on the set of trials $\mathcal{T}$ that were channel trials, $x^{\star}_{1:T}$ is the experimental perturbations presented to the participant, $\hat{\vartheta}$ is the maximum likelihood estimate of the parameters, k is the number of parameters and N is the number of data points (channel trials). The first term in the BIC penalises underfitting, whereas the second term penalises model complexity, as measured by the number of free parameters in the model. Taking the difference in BIC values for two competing models approximates twice the log of the Bayes factor. A BIC difference of greater than 4.6 nats (a Bayes factor of greater than 10) is considered to provide strong evidence in favour of the model with the lower BIC value[16]. Note that the BIC penalises model complexity more heavily than the Akaike information criterion (AIC) and corrected AIC (AICc), and hence, relative to AIC and AICc, BIC handicaps the COIN model as it has more parameters than the dual-rate model.

To perform model comparison at the group level, we calculated the group-level BIC, which is the sum of BICs over individuals[17].

4.2 Validation of model comparison: model recovery

We used the parameters from the fits of the COIN and dual-rate models to the data for each participant in the spontaneous recovery and evoked recovery experiments to generate 10 synthetic data sets for each participant from the corresponding experiment and model class (COIN and dual-rate). For a given set of parameters in a given experiment, the only source of variability in the dual-rate model across different synthetic data sets was motor noise. In contrast, for the COIN model, in addition to motor noise, sensory noise also provided a source of variability across data sets. We then fit each synthetic data set with both the COIN and dual-rate model as we did with real data. Note that for the COIN model, we reused the same synthetic data sets and fits from parameter recovery (Section 2.8.2).

For model recovery (Extended Data Fig. 2d), we examined the proportion of times the difference in BIC between the COIN and dual-rate fits favoured the true (vs. incorrect) model class that was used to generate the data.

5 Mathematical analysis of the COIN model

5.1 Spontaneous and evoked recovery

Here we develop a mathematical analysis of how the main features of spontaneous and evoked recovery emerge in the COIN model. Specifically, the main features we wish to explain are that spontaneous recovery is 1. non-monotonic, with a smooth but transient increase in adaptation, followed by decay, 2. which asymptotes (at least within the time scale of the experiment) above zero, and evoked recovery shows 3. very rapid (almost instantaneous) increase to a higher level of adaptation than spontaneous recovery, followed by monotonic decay, 4. which also asymptotes above zero.

In general, state and contextual inference in a switching state-space model, such as the COIN model, is analytically intractable. However, inference can be performed analytically under the following assumptions: (i) there are no state feedback observations (as on channel trials); (ii) the inferred parameters of the state and context transition dynamics are constant; and (iii) the number of contexts does not change. In this special case, state estimates are updated according to the state dynamics ascribed to each context:

$$\hat{x}^{(j)}_{t|t-1} = a^{(j)}\hat{x}^{(j)}_{t-1|t-2} + d^{(j)} \tag{S53}$$

and predicted context probabilities $\boldsymbol{\psi}$ are updated (independently of the states) according to the context transition matrix $\mathbf{\Pi}$:

$$\boldsymbol{\psi}_t = \boldsymbol{\psi}_{t-1}\mathbf{\Pi}, \tag{S54}$$

Assumptions (i)-(iii) are at least approximately true during the channel-trial phase of the spontaneous and evoked recovery paradigms, i.e. when our main explicanda occur. Specifically, assumption (i) is true as there is no state feedback. Assumption (ii) is approximately true as the inferred parameters governing the state and context transition dynamics are updated relatively little over the timescale relevant for spontaneous and evoked recovery late in learning. Assumption (iii) is approximately true as novel contexts tend not to be inferred when state feedback is omitted.

Based on these approximations, we simulated state and contextual inference during the channel phase of the spontaneous and evoked recovery paradigms (Extended Data Fig. 6a-c). We ran the simulations with 2 contexts using parameters $a^{(1)} = a^{(2)} = 0.95$, $d^{(1)} = -d^{(2)} = 0.0075$ and

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} & 1-\pi_{11} \\ 1-\pi_{22} & \pi_{22} \end{bmatrix}, \tag{S55}$$

where $\pi_{11} = 0.99$ and $\pi_{22} = 0.9$, reflecting the fact that nearly all transitions in the experiment are self-transitions and the context associated with P$^+$ has been experienced more often than the context associated with P$^-$.

For spontaneous recovery, on trial 1 (immediately following P$^-$), the state estimates associated with P$^+$ (context 1, red) and P$^-$ (context 2, orange) are equal but opposite (Extended Data Fig. 6a), and the context probabilities are equal (Extended Data Fig. 6b, solid lines). Hence, adaptation is initially at baseline (Extended Data Fig. 6c, solid line). Then, based on Eq. S53, the state estimates converge exponentially (at the same rate) to their steady-state values, $\hat{x}_\infty^{(j)} = d^{(j)}/(1-a^{(j)})$. In particular, for context 1, this means a monotonically decreasing decay towards a non-zero asymptote (Extended Data Fig. 6a, red) because experience in P$^+$ is compatible with a positive steady-state (which we incorporated by our choice of a positive drift rate, $d^{(j)}$, in Eq. S53). At the same time, based on Eq. S54, the predicted probabilities converge exponentially to their values under the stationary distribution, $\lim_{t\to\infty} \boldsymbol{\psi}_t$ (Extended Data Fig. 6b, solid lines). Context 1 is more probable than context 2 under the stationary distribution as P$^+$ was experienced for more trials than P$^-$ during the experiment. Hence, the predicted probability of context 1 monotonically increases (Extended Data Fig. 6b, solid red). The net result of these updates is that there is an initial rise in adaptation due to the increasing contribution of the state associated with context 1, followed by a fall in adaptation toward a non-zero baseline due to the decay of this state toward a non-zero steady-state (Extended Data Fig. 6c, solid line). Therefore, the classic non-monotonic nature of spontaneous recovery arises because the dynamics of contextual inference (responsible for the initial rise in adaptation) are faster than the dynamics of state inference (responsible for the subsequent fall in adaptation). Critically, as long as the inferred state dynamics reflect the statistics of the experiment in which, by design, the true state of each context never changes (the P$^+$ and P$^-$ perturbations are constant), the dynamics of contextual inference are bound to be faster than the dynamics of state inference, and the steady-state of adaptation (on the time scale of the experiment) to be above zero. Thus non-monotonic spontaneous recovery (feature 1) with a decay that does not reach zero (feature 2), as seen in the experimental data (Fig. 2c), is a robust feature of the COIN model. Indeed, the simulation of the full model without the approximations we introduced above for analytical tractability also shows such spontaneous recovery (Fig. 2b, bottom right, and c; also for individual participants whose data shows spontaneous recovery, Extended Data Fig. 6f) with all three main properties that our analysis here suggests are key for obtaining this result. Specifically, (a) state estimates associated with P$^+$ and P$^-$ approximately cancel at the beginning of the P^c phase (when weighted with their corresponding context probabilities) and then monotonically converge to a positive and negative steady-state, respectively (Fig. 2b, bottom left); (b) the corresponding context probabilities may be similar initially but then diverge, such that the probability associated with P$^+$ grows toward a near-one steady-state, while that associated with P$^-$ shows the opposite trend, decaying toward a near-zero baseline (Fig. 2b, top right); (c) the dynamics of contextual inference are markedly faster than those of state estimation (Fig. 2b, cf. top right and bottom left).

For evoked recovery, we assume the learner is certain they are in context 1 at the end of the second evoker (P$^+$) trial, and from then on, during the P^c trials, their contextual inferences evolve according to the same dynamics as in spontaneous recovery. For a direct comparison with spontaneous recovery, we also kept everything else (parameters, state estimates) identical to the simulation of spontaneous recovery. (This included ignoring more subtle differences in state inferences between the two paradigms; cf. Fig. 2b and d, bottom left.) Because context 1 is also much more probable than context 2 under the stationary context probabilities (see above), the context probabilities did not change much with updating from their initial values (Eq. S54), and so the probability of context 1 and context 2 remained high and low, respectively, throughout the simulation (Extended Data Fig. 6b, dashed; cf. Fig. 2d, top right). Hence, adaptation largely reflected the dynamics of the state of context 1 (Extended Data Fig. 6a, red; cf. Fig. 2d, bottom left), decaying exponentially from a level of adaptation that was necessarily higher than that reached in spontaneous recovery to a non-zero asymptote (Extended Data Fig. 6c, dashed).

5.2 Responsibility-weighted learning rate

A key prediction of the COIN model is that memory updating should depend on contextual inference (Figs. 1 and 3). This is because the COIN model assumes that only one perturbation—the perturbation associated with the current context—influences the state feedback. Hence, if the current context is known, only the memory associated with the current context should be updated after observing the state feedback:

$$\hat{x}^{(j)}_{t|t}(c_t) = \hat{x}^{(j)}_{t|t-1} + \delta_{c_t j} k^{(j)}_t e^{(j)}_t. \tag{S56}$$

However, in general, the current context is never known with certainty and so should be integrated out. After integrating out the context, the expected value of each memory update is

$$\begin{aligned} \mathbb{E}[\hat{x}^{(j)}_{t|t}(c_t)] &= \sum_{c_t} \gamma^{(c_t)}_t \hat{x}^{(j)}_{t|t}(c_t) \\ &= \hat{x}^{(j)}_{t|t-1} + \gamma^{(j)}_t k^{(j)}_t e^{(j)}_t, \end{aligned} \tag{S57}$$

where $\gamma^{(j)}_t$ denotes the responsibility of context j on trial t. Importantly, the responsibility scales the Kalman gain, producing an effective learning rate ($\gamma^{(j)}_t k^{(j)}_t$) that lies between $k^{(j)}_t$ (when $\gamma^{(j)}_t = 1$, i.e. certain that $c_t = j$) and zero (when $\gamma^{(j)}_t = 0$, i.e. certain that $c_t \neq j$). Thus contextual inference is key to Bayes-optimal memory updating (Fig. 1f).

Although the notion that error signals should be scaled by responsibilities is not unique to the COIN model[18–20], we provide experimental evidence of this computation in the memory updating experiment (Fig. 3 and Extended Data Fig. 7c-d), an achievement made possible by the recently-developed triplet assay of single-trial learning[2,21].

Here we also derive the equation shown in the 'Inference in the COIN model' subsection of the Methods for how the mean $\hat{x}^{(j)}_t$ of the predicted state distribution $p(x^{(j)}_t \mid ...)$ for context j on trial t is updated to trial $t+1$, as this provides another illustration of the responsibility-weighted nature of the learning rate. For simplicity, we assume that $p(x^{(j)}_t \mid ...)$ is Gaussian, which may be justified by invoking the central limit theorem. Under this assumption, the mean is updated across trials by combining a measurement update that incorporates the state feedback y_t (Algorithm 3) followed by a time update that simulates the state dynamics (Algorithm 2) and then integrating out the hidden context and the hidden parameters governing the state dynamics:

$$\begin{aligned} \hat{x}^{(j)}_{t+1} &= \iint \Big[a^{(j)} \left(\hat{x}^{(j)}_t + \delta_{c_t j}\, k^{(j)}_t\, e^{(j)}_t\right) + d^{(j)}\Big] p(c_t, \boldsymbol{\omega}^{(j)} \mid q_t, y_t, ...) \mathrm{d}c_t \mathrm{d}\boldsymbol{\omega}^{(j)} \\ \hat{x}^{(j)}_{t+1} &= \iint \Big[a^{(j)} \left(\hat{x}^{(j)}_t + \delta_{c_t j}\, k^{(j)}_t\, e^{(j)}_t\right) + d^{(j)}\Big] p(c_t \mid q_t, y_t, ...) p(\boldsymbol{\omega}^{(j)} \mid c_t, q_t, y_t, ...) \mathrm{d}c_t \mathrm{d}\boldsymbol{\omega}^{(j)} \\ \hat{x}^{(j)}_{t+1} &= \int \Big[a^{(j)} \left(\hat{x}^{(j)}_t + p(c_t = j \mid q_t, y_t, ...)\, k^{(j)}_t\, e^{(j)}_t\right) + d^{(j)}\Big] p(\boldsymbol{\omega}^{(j)} | c_t, q_t, y_t, ...) \mathrm{d}\boldsymbol{\omega}^{(j)} \\ \hat{x}^{(j)}_{t+1} &= \mathbb{E}_{p(a^{(j)} \mid c_t, q_t, y_t, ...)}[a^{(j)}] \left(\hat{x}^{(j)}_t + p(c_t = j \mid q_t, y_t, ...)\, k^{(j)}_t\, e^{(j)}_t\right) + \mathbb{E}_{p(d^{(j)} \mid c_t, q_t, y_t, ...)}[d^{(j)}], \end{aligned} \tag{S58}$$

where $p(c_t = j \mid q_t, y_t, ...)$ is the responsibility of context j on trial t.

5.3 Single-trial learning

Here we derive a simple and intuitive approximation to single-trial learning in the COIN model to provide insights into the memory updating experiment (Fig. 3). Single-trial learning is defined as

$$u_{t+1} - u_{t-1} = \Big(\sum_j \hat{x}^{(j)}_{t+1|t} \psi^{(j)}_{t+1}\Big) - \Big(\sum_j \hat{x}^{(j)}_{t-1|t-2} \psi^{(j)}_{t-1}\Big), \tag{S59}$$

where $\psi_{t+1}^{(j)}$ is the predicted probability of context j on trial $t+1$. To aid the derivation, we make use of the following set of simplifying assumptions:

(i) There is no decay or drift of state estimates across trials.

(ii) All state estimates are zero on the first channel trial of the triplet, which implies that errors on the exposure trial are one.

(iii) The Kalman gain is the same for all contexts.

Under these assumptions, single-trial learning can be simplified to

$$\begin{aligned} u_{t+1} - u_{t-1} &= \Big(\sum_j [\hat{x}_{t-1|t-2}^{(j)} + \gamma_t^{(j)} k_t^{(j)} e_t^{(j)}] \psi_{t+1}^{(j)} \Big) - \Big(\sum_j \hat{x}_{t-1|t-2}^{(j)} \psi_{t-1}^{(j)} \Big) \\ &= \sum_j \gamma_t^{(j)} k_t^{(j)} e_t^{(j)} \psi_{t+1}^{(j)} \\ &\propto \sum_j \gamma_t^{(j)} \psi_{t+1}^{(j)} \\ &= \boldsymbol{\gamma}_t \cdot \boldsymbol{\psi}_{t+1}. \end{aligned} \tag{S60}$$

Here $\boldsymbol{\gamma}_t$ and $\boldsymbol{\psi}_{t+1}$ are the vectors of responsibilities and predicted probabilities, respectively. Therefore, single-trial learning is approximately proportional to the dot product of the responsibilities on the exposure trial of the triplet (which determine how much each memory is updated, see Responsibility-weighted learning rate) and the predicted probabilities on the following channel trial (which determine how much each updated memory is subsequently expressed). Intuitively, this dot product is greater when the memories that are updated more are also the ones that are subsequently expressed more. In the memory updating experiment, we confirmed that single-trial learning is indeed well approximated by this dot product (Extended Data Fig. 7c-d). Moreover, the presentation of a sensory cue on the second channel trial of each triplet allowed us to reveal the effects of differential updating of a single memory by encouraging predicted probabilities to be all-or-none. In this setting, single-trial learning is proportional to the responsibility of the memory on the exposure trial, that is, when $\psi_{t+1}^{(j)} = 1$, $u_{t+1} - u_{t-1} = \gamma_t^{(j)} k_t^{(j)}$. Again, we confirmed that single-trial learning is indeed proportional to the responsibility on the exposure trial of the context with the highest predicted probability on the subsequent channel trial (Extended Data Fig. 7c-d).

The effects of perturbations, sensory cues and local transition probabilities on single-trial learning can be explained in a unified manner using this simple dot-product metric. In the memory updating experiment (Fig. 3), $\boldsymbol{\psi}_{t+1}$ is constant across the four triplet types, as we present the same sensory cue on the channel trials of all triplets, but $\boldsymbol{\gamma}_t$ varies, as we present different combinations of perturbations and sensory cues on the exposure trials of the triplets. In contrast, in the environmental-consistency experiments (Fig. 4c and Extended Data Fig. 8), $\boldsymbol{\gamma}_t$ is constant, as the same perturbation is presented on the exposure trial of the triplets and there are no sensory cues, but $\boldsymbol{\psi}_{t+1}$ varies, as the local transition probabilities differ across the environments.

6 Mapping of COIN model components to cognitive processes

6.1 Working memory in the COIN model

A working memory task performed just before the channel-trial phase has been shown to interfere with spontaneous recovery, and in fact to create an effect that is reminiscent of evoked recovery, such that P^{+} adaptation returns immediately to a high level following P^{-}, already on the first P^{c} trial (Extended Data Fig. 9a, Ref. 22). In the dual-rate model, this effect has been attributed to a selective diminishing of the adaptation of the fast learning process[22]. We simulated the COIN model with the parameters obtained from the fit to the average spontaneous and evoked recovery data sets (also used in Fig. 2b,d). The COIN model reproduces the effect by modelling the working memory task (performed after the last P^{-} trial) as selectively abolishing the (working) memory of the context responsibilities on the last P^{-} trial (Extended Data Fig. 9b-d), while sparing the (long-term) memory of context transition (and thus stationary) probabilities. This means that on the first P^{c} trial, predicted context probabilities are based on general knowledge of how frequently different contexts are expected to be encountered in the future (i.e. the learned stationary probabilities), rather than on which contexts are likely to follow the context specifically encountered on the last trial (compare coloured circles between middle right panels of Extended Data Fig. 9c and d). Because P^{+} has been the most frequent trial type, the probability of its associated context under the stationary distribution is very high, and hence there is a strong re-expression (evoked recovery) of the memory for this context. This suggests that the belief over contexts may require working memory for maintenance.

6.2 Explicit vs. implicit learning in the COIN model

Recent studies have shown that motor learning has both explicit and implicit components which exhibit markedly different time courses[23,24]. For example, in a paradigmatic example using a visuomotor rotation task, a measure of explicit learning was obtained by asking participants to report the direction in which they planned to move prior to moving, and implicit learning was then measured as the difference between the actual direction they moved and this explicit judgement[5]. In a spontaneous recovery paradigm, explicit learning showed non-monotonic behaviour during the P^{+} phase, fast increase followed by slow decay (Extended Data Fig. 9e). In contrast, implicit learning showed slower and monotonic increase during the P^{+} phase. Due to these differences in the form of adaptation, explicit and implicit learning have been suggested to correspond to the fast and slow processes, respectively, of the dual-rate model[23]. However, this mapping is unable to account for the rapid drop and recovery of supposedly slow implicit learning seen during the subsequent P^{-} and P^{c} phases.

In order to simulate these experiments, we adapted the COIN model to account for a critical difference between visuomotor and force-field learning: visuomotor but not force-field learning (which is the primary paradigm we use to test the predictions of the COIN model in the main text) introduces a discrepancy between the hand's proprioceptive and visual locations. Due to this discrepancy, a fundamental credit-assignment problem arises[20] as to whether the observed cursor deviation is due to a perturbation on the motor system or a bias (miscalibration) in the sensory system. This was naturally captured in the COIN model by introducing a bias between the state and sensory feedback as another latent parameter in each context, which was learned together with the parameters that govern the evolution of the state in that context (Methods).

We hypothesised that participants would have explicit access to the state representing their belief about the visuomotor rotation, but that they would not have access to their sensory bias, which would reflect the implicit component of learning. This hypothesis is consistent with previous work showing strong sensory recalibration during adaptation to a visuomotor rotation. For example, after learning a visuomotor rotation with their right hand, a participant can be asked to use their non-adapted left hand to point to where they sensed their right hand was at the end of a reach[25,26]. Critically, due to sensory recalibration, participants incorrectly estimate the location of their right hand location, pointing closer to where the rotated visual feedback of the hand (cursor) was than to the actual location of their right hand. This indicates that sensory recalibration remains implicit in these experiments. Our bias parameter formalises this notion of sensory recalibration, which we thus assume remains implicit.

We simulated the COIN model with the parameters obtained from the fit to the average spontaneous recovery and evoked data sets (also used in Fig. 2b,d and Extended Data Figs. 5 and 9) plus an additional parameter representing the standard deviation of the prior on the bias (Methods). Extended Data Fig. 9g, h & i show the bias, state and predicted probability for each context. The average bias across contexts weighted by the predicted probabilities (Extended Data Fig. 9j) showed a slow monotonic increase during the P^{+} phase with a drop and recovery during the P^{-} and P^{c} phase. As hypothesised, the profile is very similar to that of the implicit component of learning (Extended Data Fig. 9e-f, light green). However, the average state across contexts (Extended Data Fig. 9k) did not show the experimentally observed characteristic overshoot of the explicit component (Extended Data Fig. 9e, dark green). Instead, examining the state of the context with the highest responsibility on the previous trial (Extended Data Fig. 9h, coloured bar in the bottom, and thin black line, also shown as dark green line in (Extended Data Fig. 9f) revealed that it had a strikingly similar time course to the explicit component of learning (Extended Data Fig. 9e & f dark green). This is because the state and the bias interact competitively within a context to account for the total state feedback, and hence as the bias estimate increases, the state estimate decreases, giving rise to the characteristic non-monotonicity. As the experimental definition of explicit and implicit components guarantees that they sum to total adaptation (see above), we also defined motor output in the model as the sum of the explicit (state of the context with the highest responsibility on the previous trial) and implicit components (Extended Data Fig. 9f, solid pink). Taken together, this version of the COIN model reproduced the important qualitative features of explicit, implicit, and total adaptation in the experiment (compare Extended Data Fig. 9e and f). (Although there were quantitative differences, e.g. in the overall speed of learning, note that all but one parameter were fit to rather different force-field learning experiments and so a quantitatively precise match could not be expected.) In particular, the different time courses of explicit versus implicit components arose naturally in the model. This is because, in the COIN model, parameters (including the bias) are assumed to be constant over the lifetime of a context, and thus their estimates are updated more slowly (Extended Data Fig. 9g, j) than those of states (Extended Data Fig. 9h, k), which can change dynamically – inherently giving rise to multiple time scales of learning. Moreover, the average bias across contexts in the COIN model (Extended Data Fig. 9j, cyan, and f, light green) also tracked the rapid drop and recovery of implicit learning during the P^{-} and P^{c} phase (Extended Data Fig. 9e, light green) that the dual-rate model cannot explain. This arises from the same contextual inference-based mechanism that also underlies other aspects of spontaneous recovery (Fig. 2). Specifically, the rapid fall in the implicit component of learning during the P^{-} phase is due to the increased expression of the associated context (Extended Data Fig. 9i, orange) that has a negative bias (Extended Data Fig. 9g, orange). On entering the P^{c} phase, there is a re-expression of the context associated with P^{+} (Extended Data Fig. 9i, red) that has a positive bias (Extended Data Fig. 9g, red).

Interestingly, in order for the COIN model to be consistent with experimental data, our definition of total adaptation in this experiment (average bias plus the state of *the most responsible context* on the previous trial) needed to be different from what would have been directly consistent with the way it is originally defined in the model (the overall predicted state feedback, here corresponding to the average bias plus the average state across contexts *weighted by the predicted probabilities*). However, this experiment was also conducted differently from the other experiments we modelled. In particular, in this paradigm, an explicit judgement was solicited at the beginning of each trial before motor output was required. We reasoned that the explicit commitment of where they will aim would determine where participants eventually aim in their motor output (measured as total adaptation), thus explaining why only the reported state (corresponding to the explicit judgement in the model) and not the average state is reflected in motor output. This is in line with previous studies showing that an explicit commitment affects subsequent decision making[27]. Moreover, this reasoning made a further prediction: in a (control) variant of the same visuomotor rotation experiment in which no explicit judgements are solicited, total adaptation should have a different time course as it now should reflect the average state not just the explicitly reported state. This did indeed seem to be the case in the data (albeit slightly, not reaching statistical significance[5]): learning of P^+ was slower and adaptation to P^- was not as complete as in the original version of the task. These differences were qualitatively reproduced by the COIN model when total adaptation was modelled as usual, using the average state across contexts weighted by the predicted probabilities (Extended Data Fig. 9e-f, dashed pink).

In summary, rather than mapping explicit and implicit learning to fast and slow processes, which only differ quantitatively, the COIN model suggests that they may map to qualitatively different components of learning: state variables and bias parameters, respectively.

7 Theories of context-dependent learning

Theories of context-dependent learning have been proposed in multiple domains of cognition, including episodic memory[28] and decision making[29–34], as well as in the domain of motor control[4,18,20,35,36]. Here we give a brief unifying overview of these models. We organise our overview along the five key design choices that any model of context-dependent learning must make (even if only implicitly): what contents to attribute to each context-specific memory, how to model context dynamics, whether to use a fixed number of memories or to allow new memories to be created, how to determine the extent to which different memories are expressed at each point in time, and how to update existing memories. We close by summarising how different models of context-dependent learning fare at accounting for experimental data in the motor domain.

7.1 Memory contents

A key design choice for any model of learning (even single-context models) is specifying what is in a memory. In general, memories store information about the environment. Bayesian models of memory formalise this notion as inference over a latent variable characterising the environment given past experiences (the observations). As more experience is gained (more observations are made), these inferences can be iteratively refined, leading to the updating of memories. A critical design choice is whether the latent variable being inferred is assumed to be static over time (a 'parameter') or time-varying (a 'state'). (Importantly, even inferences about static parameters are time varying as more experience is accrued.) This choice also remains relevant for non-Bayesian models of learning. For example, models in which memories are biased towards the recent past and/or change even in the absence of experience (e.g. due to adaptive forgetting) implicitly estimate a time-varying state. In contrast, models in which memories depend equally on all past experiences (at least within the same context) and do not change in the absence of experience implicitly estimate a static parameter.

All current models of motor learning (including the COIN model) agree that a time-varying state is critical for capturing the dynamics of motor memories and therefore for understanding motor adaptation. Hence, a time-varying state forms the basis of deterministic and probabilistic models of motor learning (regardless of whether they assume the environment consists of one or multiple contexts, see below), the so-called 'state-space models'[2–4,20,36–46]. In contrast, studies of context-dependent learning in economic decision making (reinforcement learning) have typically not considered the notion of a time-varying state[19,29–31]. In these studies, participants (and models) needed to learn context-specific reward functions that determined the optimal action for each stimulus in each context[29–31]. Bayesian models of these tasks assumed that these reward functions were static and thus weighted all observations (within the same context) equally[29,31]. In one study, a non-Bayesian algorithm was used to estimate expected rewards[30]. This algorithm was a simple delta rule with a constant learning rate: it estimated expected rewards as an exponential recency-weighted average, thus implying a generative model in which the reward function may change over time, but without making explicit assumptions about how it changes. However, even this algorithm did not update its estimates in the absence of experience in a given context, i.e. at least half the time when multiple contexts exist. Therefore, in all these studies, the reward function associated with each context was mostly static over time, and hence the only truly time-varying quantity in the environment was the context.

7.2 Context dynamics

Once the notion of multiple contexts is introduced, inferring the current context becomes critical, as memory creation, expression and updating all depend on this inference (Fig. 1, see also below). In turn, this inference depends on the context dynamics, i.e. the transition probabilities between contexts. The simplest class of models assumes uniform transition probabilities (with the potential exception of a self-transition bias that makes the 'from' context the most probable), thus implying some fixed level of context-volatility[30]. A somewhat richer class of models breaks this (near) uniformity by having transition probabilities depend on the 'to' context (thus differentiating the overall occurrence of contexts) but constraining them to be the same for each 'from' context (thus rendering transitions non-Markovian/non-local[31,34]. (Again, a self-transition bias can be added[33], which might itself change over time in varying-volatility models[29].) At the other extreme are models in which transition probabilities depend on both the 'from' and 'to' contexts, without any additional constraints[35]. While these models are very flexible, they afford no generalisation between contexts, such that learning of transition probabilities needs to start afresh in each newly encountered context. A compromise between complete uniformity and extreme flexibility is provided by hierarchical models, such as the COIN model (Extended Data Fig. 1a-b), in which transition probabilities also depend on both the 'from' and 'to' contexts but are constrained to exhibit a degree of similarity via some shared global transition probabilities. Importantly, in the limit of infinite data, hierarchical models are just as flexible as their non-hierarchical counterparts (i.e. they can learn any transition matrix). However, in the finite-data regime (the most extreme case of which is when a context is encountered for the first time), hierarchical models uniquely support well-informed inferences via generalisation.

7.3 Memory creation

Broadly speaking, context-dependent models can be split into two categories: parametric and nonparametric. There are parametric models that assume that the learner knows the true number of contexts in the environment (e.g. by fixing the number of contexts/modules in the model)[4,19,20,29,35]. These models have no notion of memory creation as the number of memories is fixed from the start. In most real-world scenarios, it is unrealistic to assume that the learner knows the true number of contexts, as this number is in general only knowable through experience. This is naturally captured by nonparametric models that allow the number of contexts in the environment (however large) to be learned from experience[30–34]. These models create a new memory whenever a novel context is inferred. However, these nonparametric models used the (non-hierarchical) Dirichlet process prior. This prior assumes that there is a single distribution of contexts in the environment (analogous to the global transition distribution in the COIN model) that does not change from one time point to the next (i.e. there is no notion of context-specific/local transition distributions). In contrast, the COIN model used a hierarchical Dirichlet process prior that allows for transition probabilities to be context-specific, yet structured (see Section 7.2).

7.4 Memory expression

To generate actions in models with multiple contexts, two different approaches have been used. In one approach, the memory of the single most probable context is expressed[30,31,36]. In a second approach (also taken by the COIN model), the memories of all contexts are expressed in proportion to their respective probabilities[19,29,33,35]. The first approach ignores uncertainty about the context and can be expected to produce suboptimal actions with respect to a task-relevant loss function[47] (e.g. the squared error between an estimate of a perturbation and its true value). The second approach uses graded context probabilities to integrate out the context with respect to the loss function, allowing the optimal action to be computed and executed. In context-dependent models of reinforcement learning, these two approaches correspond to maximising expected reward for the most probable context[30,31] vs. across all possible contexts[29], respectively.

7.5 Memory updating

For models with multiple contexts (e.g. switching state-space models[4,33,36] and volatility models[29,30]), exact Bayesian inference is often intractable, requiring approximations to be made. One computationally cheap approximation that is commonly used in models of human learning is to definitively assign trials to contexts, i.e. 'hard context assignments'[30,31,33,36]. This approximation ignores uncertainty about the context and leads to a single memory being updated on each trial. This non-Bayesian heuristic has been justified on the basis that it does not qualitatively affect model behaviour[31,33]. However, this is likely to only be true for paradigms that do not directly test how memories are updated on a single-trial basis, as we do in our memory updating experiment (Fig. 3). A more accurate, though computationally more expensive, approximation is to probabilistically assign trials to contexts, i.e. 'soft context assignments' (e.g. using Monte Carlo methods, as in the COIN model), such that multiple memories are updated on each trial[29].

While many models of motor learning directly specify parameters that control generalisation both in memory expression and updating, independent of the context[41,42,48,49], the COIN model automatically and dynamically controls generalisation by principled Bayesian inference. In particular, generalisation varies over time, across contexts, and is in general different for memory expression and updating, as controlled by the predicted probabilities and responsibilities of contexts, respectively.

7.6 Explaining motor learning phenomena

Extended Data Table 1 summarises the ability of dominant single-context and multiple-context models to explain the main data sets we have modelled. We only included state-space models in this comparison, as there is broad agreement that a time-varying state is critical to capture even some of the most basic phenomena in motor learning (see Section 7.1). Hence, models without a time-varying state (such as those employed in reinforcement learning[19,29–31]) do not represent a direct alternative to the COIN model. The models used in reinforcement learning also suffer from many of the same shortcomings as the models included in Extended Data Table 1 (see Sections 7.1 to 7.5). For historical reasons, we made an exception for the MOSAIC model, as it has been a highly influential model of motor learning and is therefore directly relevant as a basis of comparison for the COIN model.

8 Neural substrates of contextual and state inference

The prefrontal cortex (PFC) is thought to play a key role in representing contextual information[50,51] and performing hierarchical inference[52]. Therefore, we expect the PFC to be the main locus of contextual inference as performed by the COIN model. Our result suggesting that working memory – known to critically rely on PFC[53] (but cf. Ref. 54) – shares the same resources with contextual inference (Extended Data Fig. 9a-d) provides further support to this idea. Ultimately, the contextual inference signals derived from the COIN model (Fig. 1f) should be predictive of neural responses in the PFC. There are several proposals for how such inferences may be encoded in neural responses[55,56] and the COIN model provides a principled tool for adjudicating between these proposals, as the posterior distributions it computes can serve as purely behaviour-based condition- and time-resolved regressors against neural data.

Adaptation in force-field learning tasks is associated with changes in neural activities in premotor and primary motor cortices, in particular during preparatory periods[57,58]. This suggests that the neural underpinning of state inference in the COIN model may be realised by adaptively tuning the initial conditions of (pre)motor cortical dynamics[59–61]. Importantly, the COIN model predicts that these circuits should be able to simultaneously maintain and adapt multiple such initial conditions (i.e. states) corresponding to different contexts. This prediction is supported by recent recordings in monkeys learning multiple force-fields, showing that changes in neural activity between contexts are orthogonal to changes that occur within each context during adaptation[57]. Finally, of the three ways in which contextual inference modulates and extends purely state inference-based mechanisms (Fig. 1b, arrows 1-3, f_{1-3}), gain control mechanisms may be ideally suited to implement graded memory expression[62], while neuromodulatory mechanisms may underlie the graded updating of memories and the creation of new memories (perhaps controlled by cholinergic and noradrenergic signals, respectively[63]).

References

1. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
2. Herzfeld, D. J., Vaswani, P. A., Marko, M. K. & Shadmehr, R. A memory of errors in sensorimotor learning. *Science* **345**, 1349–1353 (2014).
3. Smith, M. A., Ghazizadeh, A. & Shadmehr, R. Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol.* **4**, e179 (2006).
4. Heald, J. B., Ingram, J. N., Flanagan, J. R. & Wolpert, D. M. Multiple motor memories are learned to control different points on a tool. *Nat. Hum. Behav.* **2**, 300–311 (2018).
5. McDougle, S. D., Bond, K. M. & Taylor, J. A. Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *J. Neurosci.* **35**, 9568–9579 (2015).
6. Howard, I. S., Ingram, J. N. & Wolpert, D. M. A modular planar robotic manipulandum with end-point torque control. *J. Neurosci. Methods* **181**, 199–211 (2009).
7. Milner, T. E. & Franklin, D. W. Impedance control and internal model use during the initial stage of adaptation to novel dynamics in humans. *J. Physiol.* **567**, 651–664 (2005).
8. Scheidt, R. A., Reinkensmeyer, D. J., Conditt, M. A., Rymer, W. Z. & Mussa-Ivaldi, F. A. Persistence of motor adaptation during constrained, multi-joint, arm movements. *J. Neurophysiol.* **84**, 853–862 (2000).
9. Teh, Y. W., Jordan, M. I., Beal, M. J. & Blei, D. M. Hierarchical Dirichlet processes. *J. Amer. Stat. Assoc.* **101**, 1566–1581 (2006).
10. Fox, E. B., Sudderth, E. B., Jordan, M. I. & Willsky, A. S. An HDP-HMM for systems with state persistence. In *Proc. 25th Int. Conf. Machine Learning*, 312–319 (2008).

11. Carvalho, C. M., Johannes, M. S., Lopes, H. F. & Polson, N. G. Particle learning and smoothing. *Stat. Sci.* **25**, 88–106 (2010).
12. Bernardo, J. *et al.* Particle learning for sequential Bayesian computation. *Bayesian Statistics 9* **9**, 317 (2011).
13. Acerbi, L. & Ji, W. Practical Bayesian optimization for model fitting with bayesian adaptive direct search. In *Adv. Neural Inf. Proc. Sys.*, 1836–1846 (2017).
14. Jasra, A., Holmes, C. C. & Stephens, D. A. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Stat. Sci.* 50–67 (2005).
15. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Amer. Stat. Assoc.* **90**, 773–795 (1995).
16. Jeffreys, H. *The theory of probability* (OUP Oxford, 1998).
17. Li, J., Wang, Z. J., Palmer, S. J. & McKeown, M. J. Dynamic Bayesian network modeling of fMRI: a comparison of group-analysis methods. *Neuroimage* **41**, 398–407 (2008).
18. Wolpert, D. M. & Kawato, M. Multiple paired forward and inverse models for motor control. *Neural Netw.* **11**, 1317–1329 (1998).
19. Doya, K., Samejima, K., Katagiri, K.-i. & Kawato, M. Multiple model-based reinforcement learning. *Neural Comput.* **14**, 1347–1369 (2002).
20. Berniker, M. & Körding, K. Estimating the sources of motor errors for adaptation and generalization. *Nat. Neurosci.* **11**, 1454–1461 (2008).
21. Gonzalez Castro, L. N., Hadjiosif, A. M., Hemphill, M. A. & Smith, M. A. Environmental consistency determines the rate of motor adaptation. *Curr. Biol.* **24**, 1050–1061 (2014).
22. Keisler, A. & Shadmehr, R. A shared resource between declarative memory and motor memory. *J. Neurosci.* **30**, 14817–14823 (2010).
23. Mcdougle, S. D., Ivry, R. B. & Taylor, J. A. Taking aim at the cognitive side of learning in sensorimotor adaptation tasks. *Trends Cogn. Sci.* **20**, 535–544 (2016).
24. Miyamoto, Y. R., Wang, S. & Smith, M. A. Implicit adaptation compensates for erratic explicit strategy in human motor learning. *Nat. Neurosci.* **23**, 443–455 (2020).
25. Izawa, J. & Shadmehr, R. Learning from sensory and reward prediction errors during motor adaptation. *PLoS Comput. Biol.* **7**, e1002012 (2011).
26. Van Beers, R. J., Wolpert, D. M. & Haggard, P. When feeling is more important than seeing in sensorimotor adaptation. *Curr. Biol.* **12**, 834–837 (2002).
27. Stocker, A. A. & Simoncelli, E. A Bayesian model of conditioned perception. *Adv. Neural Inf. Proc. Sys.* **20**, 1409–1416 (2007).
28. Cohen, N. J. & Eichenbaum, H. *Memory, amnesia, and the hippocampal system* (MIT Press, Cambridge, Mass., 1993).
29. Findling, C., Chopin, N. & Koechlin, E. Imprecise neural computations as a source of adaptive behaviour in volatile environments. *Nat. Human Behav.* **5**, 99–112 (2021).
30. Collins, A. & Koechlin, E. Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol.* **10**, e1001293 (2012).
31. Collins, A. G. E. & Frank, M. J. Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol. Rev.* **120**, 190–229 (2013).
32. Gershman, S. J., Blei, D. M. & Niv, Y. Context, learning, and extinction. *Psychol. Rev.* **117**, 197–209 (2010).
33. Gershman, S. J., Radulescu, A., Norman, K. A. & Niv, Y. Statistical computations underlying the dynamics of memory updating. *PLoS Comput. Biol.* **10**, e1003939 (2014).
34. Sanders, H., Wilson, M. A. & Gershman, S. J. Hippocampal remapping as hidden state inference. *eLife* **9**, e51140 (2020).
35. Haruno, M., Wolpert, D. M. & Kawato, M. MOSAIC model for sensorimotor learning and control. *Neural Comput.* **13**, 2201–2220 (2001).
36. Oh, Y. & Schweighofer, N. Minimizing precision-weighted sensory prediction errors via memory formation and switching in motor adaptation. *J. Neurosci.* 9237–9250 (2019).

37. Körding, K. P., Tenenbaum, J. B. & Shadmehr, R. The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nat. Neurosci.* **10**, 779–786 (2007).
38. Forano, M. & Franklin, D. W. Timescales of motor memory formation in dual-adaptation. *PLoS Comp. Biol.* **16**, e1008373 (2020).
39. Ingram, J. N., Sadeghi, M., Flanagan, J. R. & Wolpert, D. M. An error-tuned model for sensorimotor learning. *PLoS Comp. Biol.* **13**, e1005883 (2017).
40. Ingram, J. N., Flanagan, J. R. & Wolpert, D. M. Context-dependent decay of motor memories during skill acquisition. *Curr. Biol.* **23**, 1107–1112 (2013).
41. Lee, J.-Y. & Schweighofer, N. Dual adaptation supports a parallel architecture of motor memory. *J. Neurosci.* **29**, 10396–10404 (2009).
42. Kim, S., Oh, Y. & Schweighofer, N. Between-trial forgetting due to interference and time in motor adaptation. *PLoS One* **10**, e0142963 (2015).
43. Albert, S. T. *et al.* An implicit memory of errors limits human sensorimotor adaptation. *Nat. Human Behav.* 1–15 (2021).
44. Cheng, S. & Sabes, P. N. Modeling sensorimotor learning with linear dynamical systems. *Neural Comput.* **18**, 760–793 (2006).
45. Thoroughman, K. A. & Shadmehr, R. Learning of action through adaptive combination of motor primitives. *Nature* **407**, 742–747 (2000).
46. Donchin, O., Francis, J. T. & Shadmehr, R. Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: theory and experiments in human motor control. *J. Neurosci.* **23**, 9032–9045 (2003).
47. Körding, K. P. & Wolpert, D. M. The loss function of sensorimotor learning. *Proc. Nat. Acad. Sci.* **101**, 9839–9842 (2004).
48. Poggio, T. & Bizzi, E. Generalization in vision and motor control. *Nature* **431**, 768–774 (2004).
49. Sadeghi, M., Ingram, J. N. & Wolpert, D. M. Adaptive coupling influences generalization of sensorimotor learning. *PLoS One* **13**, e0207482 (2018).
50. Donoso, M., Collins, A. G. & Koechlin, E. Foundations of human reasoning in the prefrontal cortex. *Science* **344**, 1481–1486 (2014).
51. Botvinick, M. M. Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci.* **12**, 201–208 (2008).
52. Koechlin, E. Prefrontal executive function and adaptive behavior in complex environments. *Curr. Opin. Neurobiol.* **37**, 1–6 (2016).
53. Petrides, M. Frontal lobes and behaviour. *Curr. Opin. Neurobiol.* **4**, 207–211 (1994).
54. Shallice, T. & Cipolotti, L. The prefrontal cortex and neurological impairments of active thought. *Annu. Rev. Psychol.* **69**, 157–180 (2018).
55. Aitchison, L. & Lengyel, M. With or without you: predictive coding and Bayesian inference in the brain. *Curr. Opin. Neurobiol.* **46**, 219–227 (2017).
56. Pouget, A., Beck, J., Ma, W. & Latham, P. Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* **16**, 1170–1178 (2013).
57. Sun, X. *et al.* Skill-specific changes in cortical preparatory activity during motor learning. *bioRxiv* (2020).
58. Perich, M. G., Gallego, J. A. & Miller, L. E. A neural population mechanism for rapid learning. *Neuron* **100**, 964–976 (2018).
59. Vyas, S., O'shea, D., Ryu, S. & Shenoy, K. Causal role of motor preparation during error-driven learning. *Neuron* **106**, 329–339 (2020).
60. Kao, T.-C., Sadabadi, M. S. & Hennequin, G. Optimal anticipatory control as a theory of motor preparation: a thalamo-cortical circuit model. *Neuron* **109**, 1567–1581 (2021).
61. Logiaco, L. & Escola, G. S. Thalamocortical motor circuit insights for more robust hierarchical control of complex sequences. *arXiv preprint arXiv:2006.13332* (2020).
62. Stroud, J. P., Porter, M. A., Hennequin, G. & Vogels, T. P. Motor primitives in space and time via targeted gain modulation in cortical networks. *Nat. Neurosci.* **21**, 1774–1783 (2018).

63. Yu, A. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46**, 681–692 (2005).