

Metamers of Bayesian computation

Hansem Sohn¹, Mehrdad Jazayeri^{1,2}

¹ McGovern Institute for Brain Research,

² Department of Brain and Cognitive Sciences,
Massachusetts Institute of Technology, Cambridge, MA, USA

Correspondence

Mehrdad Jazayeri, Ph.D.

Robert A. Swanson Career Development Professor
Assistant Professor, Department of Brain and Cognitive Sciences
Investigator, McGovern Institute for Brain Research
Investigator, Center for Sensorimotor Neural Engineering
MIT 46-6041

43 Vassar Street

Cambridge, MA 02139, USA

Phone: 617-715-5418

Fax: 617-253-5659

Email: mjaz@mit.edu

Acknowledgments

H.S. is supported by the Center for Sensorimotor Neural Engineering. M.J. is supported by NIH (NINDS-NS078127), the Sloan Foundation, the Klingenstein Foundation, the Simons Foundation, the McKnight Foundation, the Center for Sensorimotor Neural Engineering, and the McGovern Institute.

Author contributions

H.S. and M.J. conceived the experiments. H.S. collected and analyzed the data. M.J. supervised the project. H.S. and M.J. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

Data collected and used in this study will be made publicly available at <https://jazlab.org/resources/>.

Code availability

Code used in this study will be made publicly available at <https://jazlab.org/resources/>.

Abstract

There are two sharply debated views on how humans make decisions under uncertainty. Bayesian decision theory posits that humans optimize their behavior by establishing and integrating internal models of past sensory experiences (priors) and decision outcomes (cost functions). An alternative model-free hypothesis posits that decisions are optimized through trial and error without explicit internal models for priors and cost functions. To distinguish between these possibilities, we introduce a novel paradigm that probes sensitivity of humans to transitions between prior-cost pairs that demand the same optimal policy (metamers) but distinct internal models. We demonstrate the utility of our approach in two experiments that were classically explained by model-based Bayesian theory. Our approach validates the model-based strategy in an interval timing task but not in a visuomotor rotation task. More generally, our work provides a domain-general approach for testing the circumstances under which humans implement model-based Bayesian computations.

Introduction

Our daily life is filled with uncertain scenarios in which we have to evaluate different sources of information, infer the state of the world, and decide between different courses of action. For example, in the wake of COVID-19, we have to gather information about our surroundings to decide whether and when to eat out. This decision, like almost every other decision we make, depends on our prior beliefs (e.g., how probable it is to contract the virus), the present situation we face (e.g., how crowded the restaurant is), and the potential outcomes we expect (e.g., how bad it would be to get sick). Characterizing how humans represent and integrate these different sources of information is of utmost importance both for a deeper understanding of the nature of human cognition and for evaluating the conditions in which the decision process may go awry.

Bayesian decision theory (BDT) provides a normative framework for how to make optimal decisions under uncertainty¹⁻⁴. Imagine an agent whose decision has to be adjusted based on the state of the environment. According to BDT, the optimal decision policy can be computed from three sources of information (Figure 1a): the prior probability distribution of the underlying state, the likelihood function of the state derived from noisy sensory evidence, and a cost (or reward) function that specifies possible decision outcomes. Optimal integration of these ingredients would enable the agent to maximize its expected reward (or minimize expected cost).

Remarkably, BDT can capture human behavior in a variety of domains including perception⁵⁻¹⁰, sensorimotor function¹¹⁻¹⁴, multimodal integration¹⁵⁻¹⁸, and high-level cognitive function¹⁹⁻²¹. Based on the remarkable success of BDT, it has been proposed that the brain performs Bayesian computations by explicitly representing the sensory likelihood, prior knowledge, and cost function²²⁻²⁶.

However, the notion of the Bayesian brain is fiercely debated²⁷⁻²⁹. It is argued that the success of Bayesian models is unsurprising given the degrees of freedom researchers have in choosing the prior distribution, likelihood function, and cost function. The crux of the disagreement is about the value of formulating the optimization process in terms of a specific prior distribution and a specific cost function given that learning these components is not essential for learning the optimal policy^{22,30,31}. A perfectly reasonable model-free alternative (Figure 1a) is that humans use trial-by-trial observations to incrementally arrive at the optimal solution without explicit reliance on the prior distribution and/or cost function³²⁻³⁶. Moreover, from a theoretical perspective, the choice of the prior and cost function is not unique. For example, an optimal agent may choose an option more frequently either because it is more probable or because it is more rewarding. More generally, in decision-making tasks, there are usually numerous pairs of priors and cost functions that, when combined, could lead to indistinguishable decision policies (Figure 1b). Analogous to the notion of metamers in perception³⁷⁻³⁹, we will refer to such pairs of priors and cost functions as prior-cost metamers. Because of the existence of such metamers, it remains an important and unresolved question whether decisions are made based on independently learned priors and cost functions.

Here we turn the problem of prior-cost metamers on its head to develop a general experimental strategy to test whether human decisions rely on independently learned priors and cost functions. The key idea behind our approach is to ask whether human behavior in a decision making task exhibits signs of relearning when we covertly switch from one prior-cost pair to another pair that is associated with exactly the same optimal decision policy (Figure 1c). According to BDT, optimal behavior depends on having

learned the prior and cost function independently. Under this model-based hypothesis, changing to a new pair would lead to a transient change in decision policy until the observer relearns the new pair. Alternatively, if the optimal behavior were to emerge through trial-and-error without learning the prior and cost, then the observer should show no sensitivity to the switch and the behavior should remain optimal. We applied this approach to two tasks – a time-interval reproduction task^{40,41} and a visuomotor rotation task⁴² – both of which were classically explained in terms of BDT. Our results substantiated the role of independently learned priors and cost functions in the timing task, but not in the visuomotor rotation task. Accordingly, future behavioral studies can take advantage of our approach based on metamers to rule out the possibility of non-Bayesian strategies in Bayesian-looking behaviors.

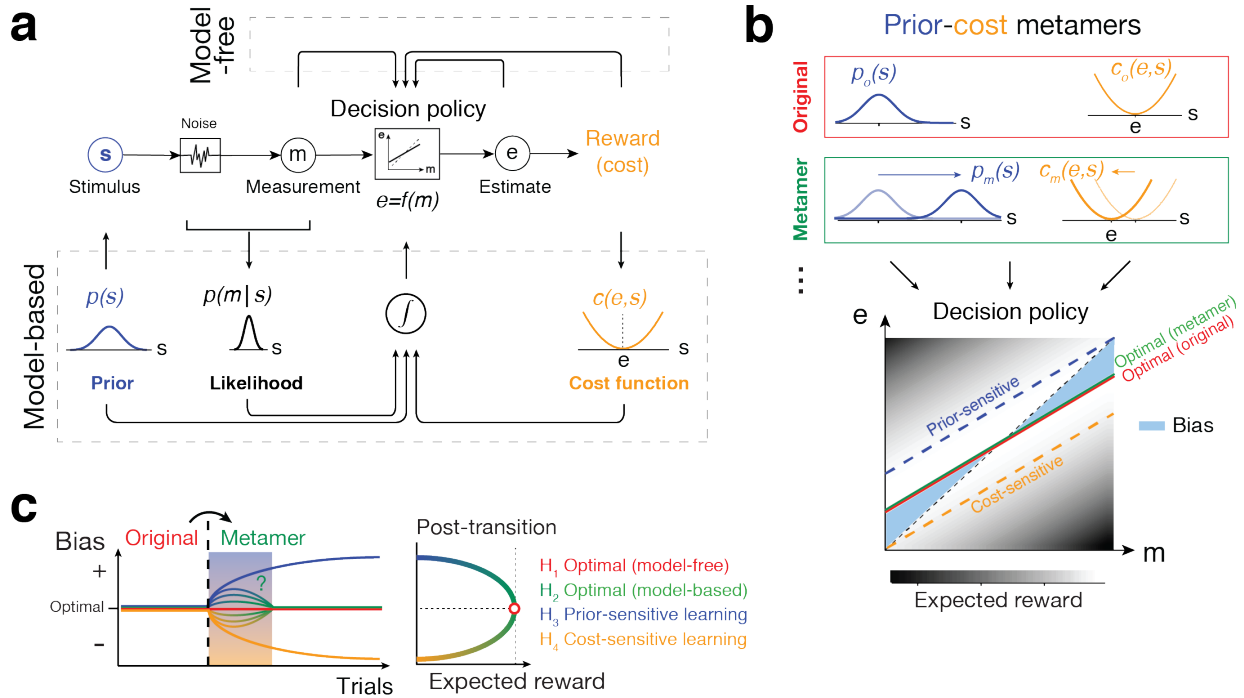


Figure 1. Testing Bayesian models of behavior using prior-cost metamers.

a. Ideal observer model. An agent generates an optimal estimate (e) of stimulus (s) based on a noisy measurement (m). To do so, the agent must compute the optimal decision policy (rectangular box; $e=f(m)$) that maximizes expected reward. The decision policy can be computed using either a model-free (top) or model-based (bottom) learning strategy. In model-free learning, the policy is optimized through trial and error based on measurements and decision outcomes (arrows on top). In model-based learning, the agent derives the optimal policy by forming internal models for the stimulus prior probability, $p(s)$, the likelihood of the stimulus after measurement, $p(m|s)$, and the underlying cost function, $c(e,s)$.

b. Prior-cost metamers are different $p(s)$ and $c(e,s)$ pairs that lead to the same decision policy. Red box: An original pair (subscript o) showing a gaussian prior, $p_o(s)$, a quadratic cost function, $c_o(e,s)$. Green box: A metamer pair (subscript m) whose gaussian prior, $p_m(s)$, and quadratic cost function, $c_m(a,s)$, are suitably shifted to the right and left, respectively. Bottom: The optimal decision policy ($e=f(m)$) associated with both the original (red) and metamer (green) conditions is shown as a line whose slope is less than the unity line (black dashed line; unbiased). The colored dashed lines show suboptimal policies associated with an agent that is only sensitive to the change in prior (blue) or only sensitive to the change in the cost function (orange). The policy is overlaid on a gray-scale map that shows expected reward for various mappings of m to e .

c. Hypothetical behavioral results after an uncued transition (vertical dashed line) from the original pair to its metamer. H_1 (model free): After the transition, the agent continues to use the optimal policy associated with the original condition. Since this is also the optimal policy for the metamer, response biases do not change and expected reward remains at the maximum level (right). H_2 (model-based): Immediately after the transition, the agent has to update its internal model for the new prior and cost function. This relearning phase causes a transient deviation from the optimal bias (curved lines after the switch, left) and reduces expected reward (right). After learning, the behavior becomes optimal since the optimal policy for the metamer is the same as the original. H_3 (prior-sensitive): The agent only learns the new prior, which leads to a suboptimal behavior and reduced expected reward. H_4 (cost-sensitive): Same as H_3 for an agent that only learns the new cost function.

Results

Although we can use Newton's Laws to explain how an apple falls from a tree, we would not conclude that the apple 'implements' Newton's Laws. In the same vein, the fact that we can explain a person's behavior in terms of the Bayesian decision theory (BDT) does not necessarily mean that their brain relies on explicit representations of priors and cost functions. Here, we apply a new experimental approach to investigate whether Bayesian computations rely on explicit knowledge about priors and cost functions. We explore this question in the context of two behavioral tasks, a time-interval reproduction task and a visuomotor rotation task. Previous work has shown that human behavior in both tasks is nearly optimal and can be explained in terms of integrating priors and cost functions^{40,42}. Here, we test whether optimal performance in these tasks can indeed be attributed to explicit reliance on priors and cost functions.

Our experimental approach for both tasks is the same. We start the experiment with a specific choice of prior and cost function; after performance saturates, we covertly switch to a prior-cost metamer associated with the same optimal policy. The key question we ask is whether the transient behavior immediately after the switch reflects any sign of relearning the new prior-cost pair. We use the transient behavior to distinguish between two hypotheses. Under one hypothesis (H_1 , Figure 1c), the optimal behavior emerges through trial-and-error without explicit learning of the prior and cost. We will refer to this strategy as model-free learning since the optimality does not rely on an explicit model of the prior and/or cost. This hypothesis predicts that behavior shows no sensitivity to the switch and remains optimal because the optimal policy is the same as the one associated with the originally learned prior-cost pair. Under another hypothesis (H_2), optimal behavior depends on learning the prior and cost as prescribed by the BDT. We will refer to this strategy as model-based learning since the agent has to build an internal model for the prior and cost. This hypothesis predicts that the behavior transiently deviates from optimality while the new prior and cost are being learned, but the asymptotic behavior would be optimal, as H_1 . For comparison, we will also test the two additional hypotheses, one in which the behavior is assumed to be solely sensitive to the prior (H_3), and one in which the behavior is assumed to be solely sensitive to the cost function (H_4). These hypotheses provide an informative benchmark to examine whether participants exhibit differential sensitivity to the prior or cost function.

Time-interval reproduction task: Ready-Set-Go (RSG)

In the RSG task (Figure 2a), participants measure a time interval between the first two beats of an isochronous rhythm ('Ready' and 'Set' flashes) and are asked to press a button at the expected time of the third omitted beat ('Go'). We refer to the sample interval between Ready and Set as t_s , and the production interval between Set and Go as t_p . We evaluated performance in RSG by what we refer to as a 'metronomic function', which quantifies t_p as a function of t_s . On each trial, t_s is sampled from a prior probability distribution and participants receive graded numeric feedback as reward (R) depending on the absolute error between t_p and t_s . We set the value of R according to a truncated quadratic function with a maximum of 100 and a minimum of 0 (Figure 2c). We will use the terms reward and cost function interchangeably to refer to the functional relationship between R and error ($t_p - t_s$).

Before analyzing the participants' behavior, we characterized the predictions of various hypotheses regarding behavior in the context of two metameric pairs. Under both H_1 and H_2 , the behavior is expected to be asymptotically optimal (i.e., after learning). We determined the form of the optimal policy by characterizing the behavior of an ideal observer performing the RSG task. An ideal observer integrates

the likelihood function associated with the measured interval (t_m) with the prior and the cost function to derive an optimal estimate (t_e) that maximizes expected reward. This Bayes-optimal integration manifests as a nonlinear relationship between t_e and t_m (Figure 2c, bottom right). Accordingly, the t_p values for a participant that employs an optimal strategy should exhibit characteristic biases toward the mean of the prior. At first glance, the bias in the optimal policy may seem counter-intuitive. However, biasing responses in this way reduces the variance of responses such that the performance improvement due to reduced variance is larger than the performance drop due to the addition of biases. We also predicted the behavior for the prior-sensitive (H_3) and cost-sensitive (H_4) hypotheses after the switch to its metamer. According to H_3 , t_p values should exhibit an overall positive bias toward longer intervals. In contrast, under H_4 , t_p values should exhibit an overall negative bias. Moreover, since both hypotheses are suboptimal, they predict an overall drop in expected reward.

Next, we collected data from 11 participants performing the RSG task in the context of two distinct metameric prior-cost pairs across 5 daily sessions (Figure 2b). In the first session, participants performed the task under the 'original' pair, which consisted of a Gaussian prior distribution and a cost function that was centered at zero when $t_p = t_s$ (Figure 2c; red). For the original pair, participants' t_p increased with t_s , and exhibited systematic biases toward the mean of the prior (Figure 2d, red). Similar to numerous previous studies^{40,41,43-46}, this behavior was consistent with predictions of the ideal observer model (Supplementary Figure 1).

Our goal for the subsequent four sessions was to test participants' behavior after a covert switch between the original pair and its 'metamer'. To design the metamer (green in Figure 2c), we used the same truncated quadratic form for the cost function but shifted it by -150ms (Δt) such that the maximum value of R was now associated with responding 150 ms earlier than the third beat ($t_p = t_s - 150\text{ms}$). Next, we designed the metameric prior. Intuitively, the new prior has to be shifted in the positive direction to counter the negative shift in the cost function and lead to the same decision policy. However, to achieve a perfect metamer, the new prior cannot be Gaussian (see Methods). We modeled the new prior as a Gaussian Mixture Model (GMM) whose parameters were adjusted such that the integration of the GMM and shifted cost function produced the same optimal policy. Note that the exact form of the GMM depends on measurement noise level and was therefore customized for each participant independently (see Methods). For each subject, we verified that the optimal policy for the original and metamer were indeed the same (Supplementary Figure 2).

Participants' behavior in the context of the metamer showed similar biases toward the mean (Figure 2d, green; see Supplementary Figure 2 for all subjects) and was asymptotically matched to that of the ideal observer (Figure 2e, green; $p=0.137$, t test for equal bias between data and optimal model). In other words, participants' behavior during both the original and its metamer was captured by the same optimal policy. This finding is consistent with both the model-free (H_1) and model-based (H_2) optimal strategy.

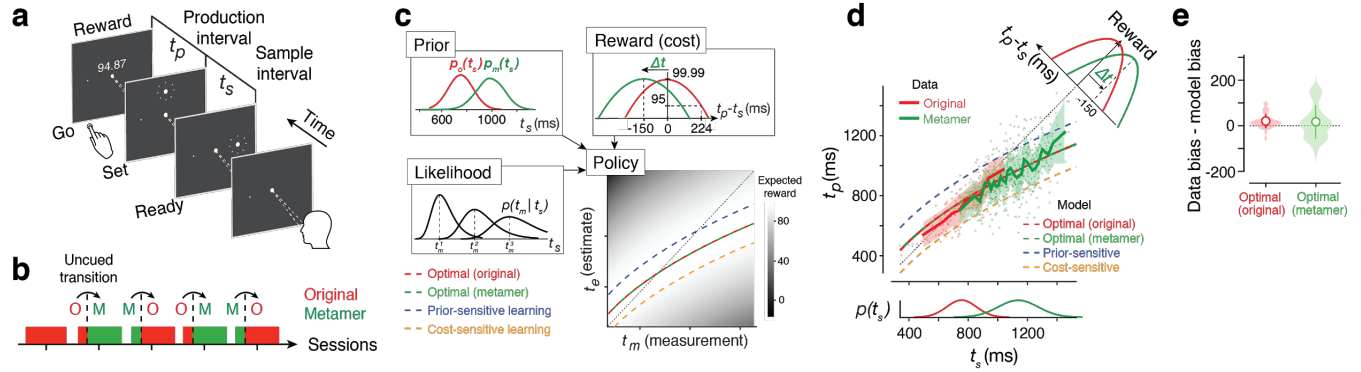


Figure 2. Prior-cost metamers in the Ready-Set-Go task.

a. Time-interval reproduction task (“Ready-Set-Go”). While looking at a central spot, participants measure a sample interval (t_s) demarcated by two flashes (Ready and Set), and initiate a delayed button press (Go) to produce a matching interval (t_p) immediately after Set. At the end of each trial, participants received a numerical score whose value was determined by a cost/reward function (see below).

b. Experimental sessions. In the first session that served as a baseline, participants performed the task with the original prior-cost set (‘O’). In the subsequent 4 sessions, we introduced an uncued transition between the original and metamer (‘M’) in an alternating fashion. In these sessions, the pre- and post-transition included ~ 170 , and ~ 500 trials, respectively. In all panels we use red for the original condition and green for the metamer.

c. Prior-cost metamers. Top-left: the original prior, $p_o(t_s)$, is a Gaussian distribution (mean: 750 ms, SD: 144 ms) and its metamer, $p_m(t_s)$, is a mixture of Gaussian distribution (see Methods). Top-right: reward (or cost) function is an inverted quadratic function of the error, $t_p - t_s$, that is truncated to have only positive values. The original cost function is centered at zero error, and the metamer is the same function shifted by Δt . Bottom-left: The likelihood function for t_s based on noisy measurement, t_m , denoted $p(t_m|t_s)$. The likelihood function is asymmetric and broader for longer t_s because measurement noise is assumed to scale with t_s . Bottom-right: The optimal policy function that prescribes how an ideal observer should derive an estimate, t_e , from t_m . By design, the optimal policy for the original (H_1) and metamer (H_2) are identical. The plot also shows suboptimal policies for a prior-sensitive (blue; H_3) and cost-sensitive (orange; H_4) overlaid on a grayscale map showing average expected reward for different mapping of t_m to t_e . When the original pair switches to its metamer, H_3 and H_4 predict positive and negative offsets, respectively, relative to the optimal policy.

d. Data from a representative session showing t_p as a function of t_s (dots: individual trials, solid lines: running average across trials; shaded area: SEM). Prediction from H_1 to H_4 are shown in the same format as panel c. Below: The original and metamer priors. Top-right: The original and metamer cost functions.

e. Bias in the original and metamer contexts. Violin plot showing the difference between the observed and optimal bias across all participants and all sessions with transitions (N=40, shaded region: distribution, open circles: grand averages; error bars: standard deviation). Bias was computed as the root mean squared difference between running average of t_p and t_s .

Our main interest, however, is to distinguish between the model-free (H_1) and model-based (H_2) strategies. Recall that any transient deviation from the behavior of an ideal observer after the switch would provide evidence against the model-free strategy of using the original policy (no re-learning). Therefore, we estimated the log-probability of behavioral data under the ideal-observer model before, immediately after, and long after the covert switch. As expected, the magnitude of bias across the participants was nearly matched to that of an ideal observer before the switch (Figure 3a, right inset). After the transition, the probability of data under the ideal observer model dropped sharply and transiently (Figure 3a, left), and within approximately 50 trials, moved back to the pre-transition level (nonparametric Friedman test for effect of transition on mean $p(\text{optimal policy})$ across [-25 0], [1 25], [26 50] trials with respect to the transition, $p=0.020$; post-hoc signed rank test for 25 trials before versus after the transition, $p=0.081$; post-hoc signed rank test for 1-25 trials versus 26-50 trials after the transition, $p=0.003$). It is possible that the transient deviation from optimality was due to the fact that participants experienced new t_s values after the prior was switched. We ruled out this possibility by analyzing only the trials whose t_s values overlapped with the distribution of t_s before the transition (Supplementary figure 3). Note that the transient deviation from the ideal observer behavior was small in comparison to what is expected from a purely prior-sensitive (H_3 ; Figure 3b) or cost-sensitive (H_4 ; Figure 3c) strategy. Together, these results reject H_1 , H_3 and H_4 , but not H_2 , suggesting that participants were indeed sensitive to both the prior and cost function despite the metameric relationship between the prior-cost pairs before and after the switch.

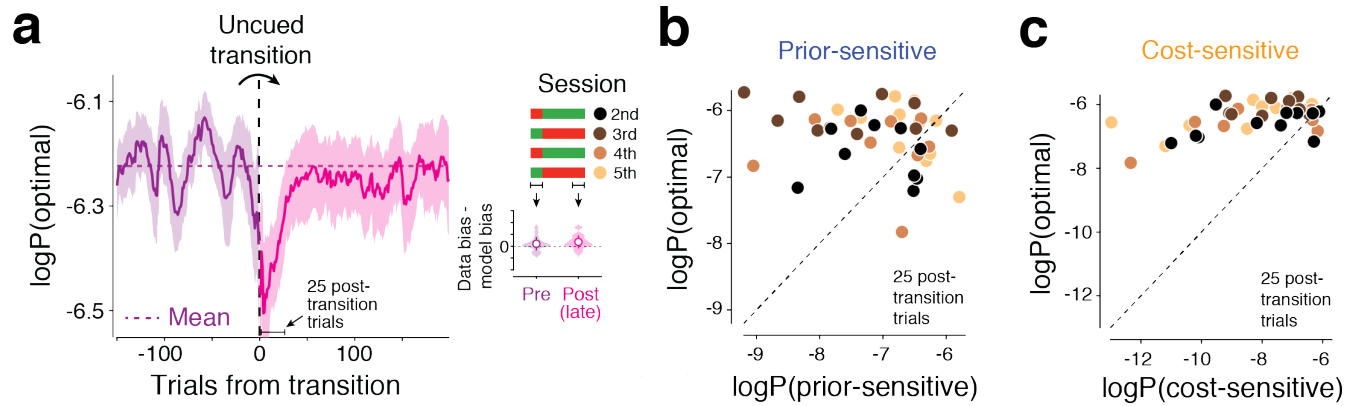


Figure 3. Behavior immediately after the switch to a new prior-cost condition in the RSG task

a. Log probability of the optimal policy, denoted ' $\log P(\text{optimal})$ ' under the data across a few hundred trials before and after the transitions (vertical dashed line) averaged across subjects and sessions ($N=40$, shaded area: SEM, purple: pre-transition, magenta: post-transition; horizontal dotted line: log probability of the optimal policy averaged across 100 pre-transition trials). Data from individual trials are smoothed using a causal Gaussian kernel (SD: 10 trials), separately for the pre- and post-transition. Twenty five post-transition trials are also highlighted as a window of interest for later analyses. Legend illustrates the alternating transitions between the original (red) and metamer (green) across 4 sessions. Right inset: the average difference between observed and optimal bias in the pre- and post-transition epochs plotted in the same format as Figure 2e. Pre-transition includes all trials before transition. Post-transition includes the last 200 trials (to avoid misestimation due to the transient).

b. Comparison of the optimal policy, $\log P(\text{optimal})$, with the suboptimal prior-sensitive policy, denoted $\log P(\text{prior-sensitive})$, based on the data in the 25 post-transition trials across participants and sessions ($N=40$; see legend in **a**).

c. Same as **b** for the cost-sensitive model.

Learning dynamics of the prior-cost metamer

So far, we have established that participants were sensitive to transitions between prior-cost metamers, which suggests an underlying dynamic process for learning the new prior and cost function. To characterize this dynamic process, we sought to track the participants' internal estimate of the prior and cost function during the learning process. To do so, we built an observer model in which the prior and cost function could change dynamically during learning. From a statistical (fitting) perspective, parameterizing the full prior distribution and cost function throughout learning is not feasible. We therefore built a simplified model in which the profile of the prior and cost function were the same as those associated with the original pair (i.e., Gaussian and quadratic, respectively). However, we allowed the mean (μ) and standard deviation of the prior, and the shift in the cost function (Δt) in the model to be determined dynamically based on the observed behavior. We will use 'subjective prior' and 'subjective cost function' to refer to μ and Δt , respectively.

We fitted this observer model to each participant's behavior using a running window of 100 trials and tracked the fits to subjective prior and cost function from before to after the transition. Across participants, μ and Δt changed systematically and in accordance with the changes in the experimentally imposed prior and cost function (Figure 4a,b; see Supplementary figure 4 for the standard deviation of the prior). We also use Bayesian Information Criterion (BIC) to evaluate the data with respect to the model-free (H_1), model-based (H_2), prior-sensitive (H_3), and cost-sensitive (H_4) hypotheses (Figure 4c). Results provided clear evidence in support of the model-based hypothesis. Together, these analyses substantiate the presence of a dynamic learning process for the prior and cost function after the transition, and provides an estimate of the underlying time course of learning in this task.

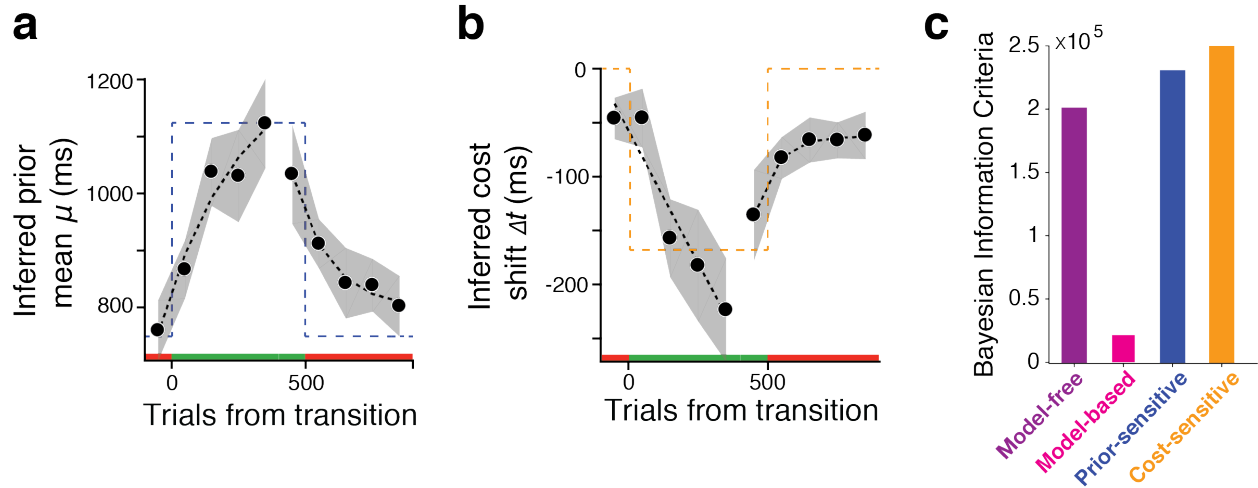


Figure 4. Time course of learning the new prior and cost function across subjects.

a,b. The inferred mean of the prior (μ) during two types of transitions between the original and metamer conditions (red and green above the abscissa). The mean was inferred from fits of an observer model to behavior in running windows of 100 trials at different time points before and after the transition (black circles: average across participants and sessions $N=40$, shades: SEM, black dotted line: exponential fit; dashed line: the experimentally imposed prior mean).

b. The inferred shift in the subjective cost function (Δt) during transitions between the original and metamer conditions shown with the same format as **a**.

c. Model comparison using Bayesian information criteria (BIC). Smaller BIC values indicate the better models.

One important observation from the learning dynamics was that the prior and cost function changed in parallel and their learning rates appeared to be comparable (Figure 4a,b). Indeed, fitting the learning curve for the prior and cost function with an exponential function led to comparable learning rates (Figure 5a; $p=0.42$, signed rank test). To gain a deeper understanding of the computational consequences of this parallel learning we performed a series of simulations in which we varied the learning rate for the prior and cost function (Figure 5f). The simulations indicated that when the prior learning was faster than the cost function, the behavior after the transition became more like the prior-sensitive model (H_3) and led to a reduction of expected reward. A higher learning rate for the cost function also reduced performance by causing the behavior to become more like the suboptimal cost-sensitive model (H_4). The best performance during transition was associated with cases when two learning rates were comparable. This result provides a normative argument that it may be beneficial to have comparable learning rates for the prior and cost.

To address the question, we performed a simulation to estimate the expected reward in a state space (Figure 5f,g) comprised of the mean of the subjective prior (μ) and the shift in the cost function (Δt). As expected, when the subjective prior mean and cost shift match the objective values (original and metamer in Figure 5f,g), the expected reward is maximum. Any other point in the space would give rise to suboptimal reward amounts. However, we unexpectedly found a continuum of the prior-cost pairs that can achieve almost the maximum reward, in between the original and metamer sets we devised. Therefore, if the subjects learned the new prior and cost in parallel – i.e., navigating the state space diagonally, the expected reward would not decrease as subjects remain in the continuous regime of the optimal reward. However, the reward would decline rapidly as subjects sequentially update their internal prior or cost function.

One intriguing conclusion from these simulations is that when prior and cost learning are suitably coordinated, the behavior may remain optimal throughout the relearning process. In other words, a participant's behavior may show no sign of learning (i.e., no deviation from ideal observer model) even while they are in the process of learning the new prior and cost function. With this consideration in mind, we analyzed the behavior of individual participants asking whether such coordinated learning for the prior and cost function was evident in their behavior (Figure 5b-e). Results revealed a diversity of learning rates for the prior and cost function ranging from relatively faster prior learning (Figure 5b), to comparable learning rates for the prior and cost function (Figure 5c,d), to faster cost learning (Figure 5e). However, there was no systematic difference between the two learning rates across participants (Figure 5a).

The diversity of learning rates for the prior and cost function across participants provides a coherent explanation for various findings that we originally found puzzling. First, it explains why we were able to infer the prior mean and the shift in the cost function across participants (Figure 4a,b). If the learning dynamics were identical across participants, we would have not been able to infer those dynamics because of the metameric relationship between the two conditions. Second, since the prior and cost learning rates were comparable across participants, the deviation from the ideal observer policy after prior-cost switches was small (Figure 3a) compared to the prior-sensitive and cost-sensitive strategies (Figure 3b,c). Third, the parallel learning enables participants to maintain a steady performance after prior-cost switches (Supplementary Figure 5).

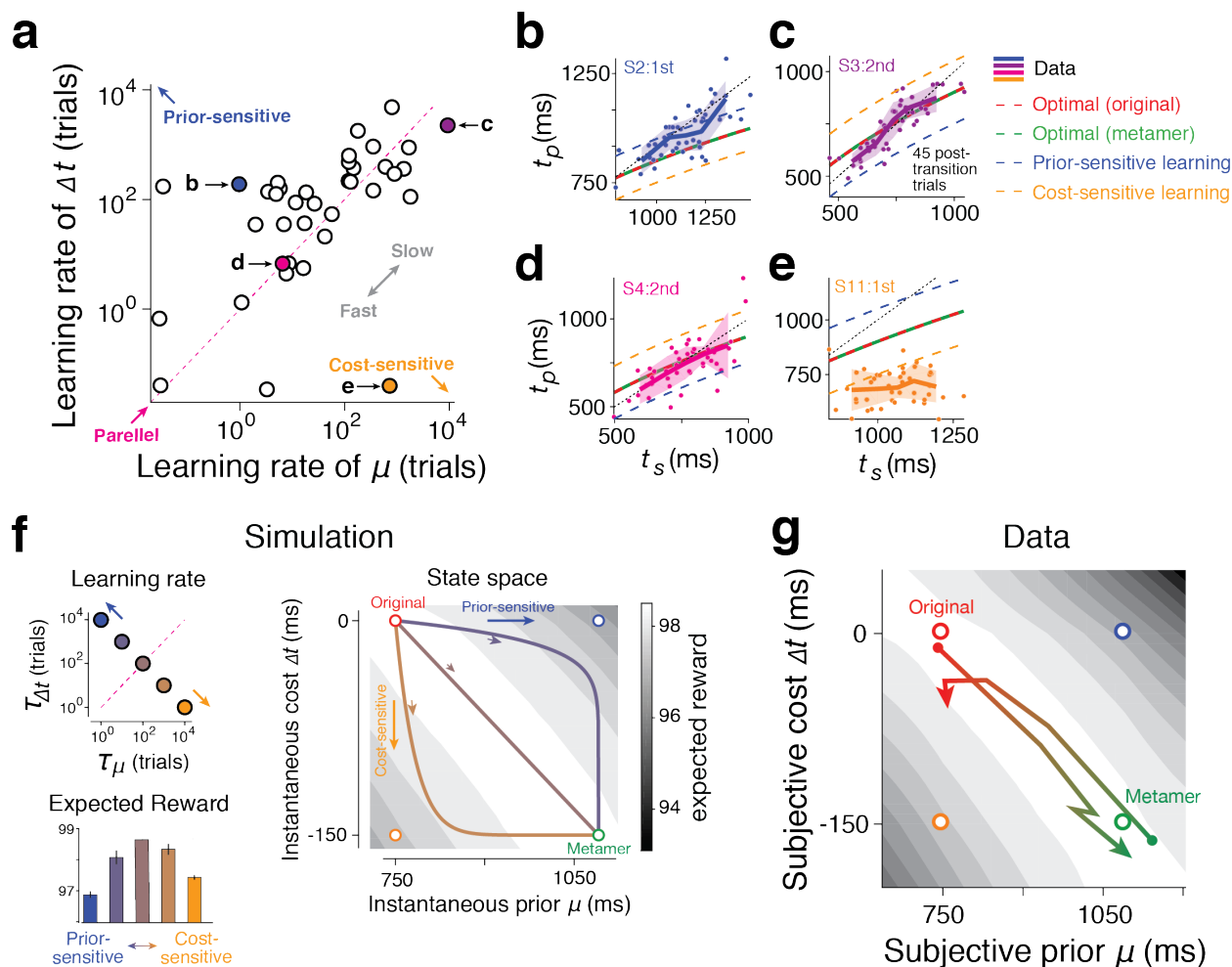


Figure 5. Parallel learning of prior and cost function and its computational consequences.

a. Learning rate for adjusting the mean of the prior (μ) and the shift in the cost function (Δt) after prior-cost transitions. We inferred the learning rates by fitting exponential curves to learning dynamics (Figure 4). For many participants, the learning rates were comparable as evident by the points near the unity line ('parallel', magenta dotted line). Points above the unity line indicate faster adjustment for μ relative to Δt , which is expected from a participant that is more prior-sensitive (blue circle). Points below the unity line indicate faster adjustment for Δt relative to μ , which is expected from a participant that is more cost-sensitive (orange circle). Two outlier data points on top-right were not included in the plot (N=40 across subjects and sessions).

b-e. Behavior of four example participants in the first 45 post-transition trials (dots: individual trials; solid line: average across bins; shaded region: SEM). The title of each plot indicates the participant (e.g., S2) as well as the session with transitions (e.g. '2nd'), and the colors correspond to the specific points in panel **a**. For each example, the predictions from various optimal and suboptimal models are also shown (legend).

f. Simulation with varying learning rates. Top-left: We selected 5 pairs of the learning rates for μ and Δt , indexed by the corresponding time constants, τ_μ , and $\tau_{\Delta t}$, respectively. The pairs ranged from faster prior learning (blue) to parallel learning (brown) to faster cost learning (orange). Right: Learning trajectory of μ and Δt for three example points in the top-left panel (with the corresponding colors) during the transition from the original prior-cost set (red circle) to its metamer (green circle). Faster prior learning first moves the state toward the prior-sensitive model (open blue circle) and then toward the end point. Faster cost learning first moves the state toward the cost-sensitive model (open orange circle). When the learning rates are identical (brown), the state moves directly from the original to the metamer. The grayscale colormap shows expected reward as a function of μ

and Δt (Equation 19 in Methods). Bottom-left: Average reward along the learning trajectory for the points in the top-left panel. Results correspond to the averages of 500 simulated trials (error bar: SEM).

g. Learning trajectories across participants based on average learning rates in panel **a** shown separately while transitioning from original to metamer (red to green), and from metamer to original (green to red). Note that, on average, the expected reward (grayscale colormap, same as panel (f)) remains high.

Visuomotor rotation (VMR) task

In the VMR task (Figure 6a), participants use a manipulandum to move a cursor from the center of a visible ring to the remembered position of a target flashed briefly on the circumference of that ring. As soon as the cursor begins to move, we make the cursor invisible and change its angular position relative to the angular position of the hand by a rotation angle, x_s . While moving, the cursor reappears briefly when it is midway, which provides information about x_s to the participants. On each trial, x_s was sampled from a Gaussian prior distribution (Figure 6c), and a numerical feedback, R , was provided depending on the error between x_s and the corresponding correction, x_p , which we defined as the angle between the hand position and the target on the ring. We set the value of R according to a truncated quadratic function with a maximum of 100 and a minimum of 0 (Figure 6c).

Similar to the RSG task, we characterized the predictions of various hypotheses regarding behavior in the context of two metamer pairs. Under the model-free (H_1) and model-based (H_2) Bayesian strategies, the behavior is expected to be asymptotically optimal (i.e., after learning). We determined the form of the optimal policy by characterizing the behavior of an ideal observer that integrates the likelihood function based on a noisy measurement, x_m with the prior and the cost function to derive an optimal estimate, x_e , that maximizes expected reward (Figure 6c). With a Gaussian likelihood function, a Gaussian prior and a quadratic cost function, the optimal policy is to map x_m to x_e linearly with a bias toward the mean (i.e., slope of less than 1; see Methods). Under the prior-sensitive (H_3) and cost-sensitive (H_4) hypotheses, x_p values would exhibit an overall positive and negative bias, respectively (Figure 6c), and would lead to lower reward than expected under the optimal policy. Note that linearity of the optimal policy in the VMR task is due to the relatively simple form of the likelihood function (Gaussian with a fixed standard deviation) compared to the RSG task in which variability scales with the base interval.

We collected data from 10 of the participants who also performed the RSG task. Our experimental procedure was the same as in RSG: an original prior-cost pair in the first session, and alternations between the original and its metamer in the subsequent 4 sessions (Figure 6b). The original prior was centered at -12 deg, and the original cost function was centered at zero. Participants' behavior exhibited biases toward the mean as expected by the optimal policy (Figure 6d for representative data; see Supplementary figure 6 for all participants; Figure 6e, red, $p=0.561$, t test, Null: equal bias between data and optimal model).

Next, we designed the prior-cost metamer (Figure 6c). For the prior, we shifted the Gaussian distribution by 10 deg so that the new mean was at -2 deg. For the cost function, we derived the appropriate shift in the cost function (Δt) needed to create a metamer pair (i.e., matching optimal policy). Note that the shift in the cost function had to be customized separately for each participant depending on the measurement noise fits in the first session. Participants' behavior in the context of the metamer showed similar biases toward the mean (Figure 6d, green; Supplementary Figure 6 for all participants) and was asymptotically matched to that of the ideal observer (Figure 2e, green; $p=0.052$, t test for equal bias for data and optimal model). In other words, participants used the same optimal policy for both the original and its metamer, consistent with both the model-free (H_1) and model-based (H_2) strategy.

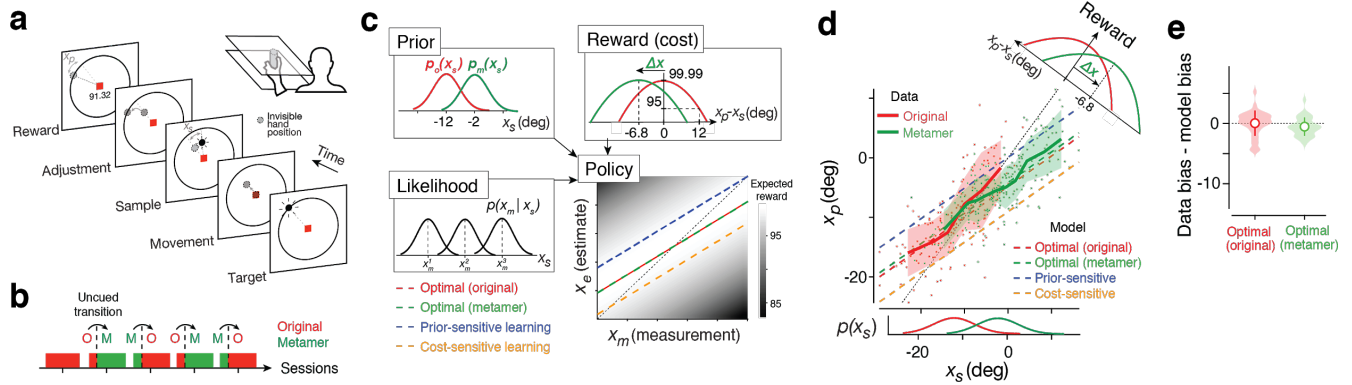


Figure 6. Prior-cost metamers in the visuomotor rotation task.

- a.** Visuomotor rotation task. Participants move a manipulandum (vertical cylinder) to maneuver a cursor on a horizontal display (top-right) from the center of the screen (red square) to a target flashed briefly on the circumference of a visible ring ('Target': black circle on the ring). The moving cursor is occluded while moving except from a brief reappearance halfway along the path ('Sample': black circle). The cursor position is covertly rotated by an angle (x_s) relative to the position of the hand (gray circle). Participants have to use x_s and apply a matching counterrotation (x_p) to bring the cursor to the target position ('Adjustment'). At the end of each trial, participants received a numerical score whose value was determined by a cost/reward function (see below).
- b.** Experimental sessions shown in the same format as Figure 2b.
- c.** Prior-cost metamers shown in the same format as in Figure 2c. Top-left: the original prior, $p_o(x_s)$, is a Gaussian distribution (mean: -12 deg, SD: 7.5 deg) and its metamer, $p_m(x_s)$, is also a Gaussian distribution shifted by 10 deg (see Methods). Top-right: reward (or cost) function is an inverted quadratic function of the error, $x_p - x_s$, that is truncated to have only positive values. The original cost function is centered at zero error, and the metamer is the same function shifted by Δt . Bottom-left: The likelihood function for x_s based on noisy measurement, x_m , denoted $p(x_m|x_s)$. Bottom-right: The optimal policy function that prescribes how an ideal observer should derive an estimate, x_e , from x_m . By design, the optimal policy for the original (H_1) and metamer (H_2) are identical. The plot also shows suboptimal policies for a prior-sensitive (blue; H_3) and cost-sensitive (orange; H_4) overlaid on a grayscale map showing average expected reward for different mapping of x_m to x_e . H_3 and H_4 predict positive and negative offsets, respectively, relative to the optimal policy.
- d.** Data from a representative session showing x_p as a function of x_s shown in the same format as in Figure 2d.
- e.** Bias in the original and metamer contexts ($N=36$) in the same format as in Figure 2e.

To distinguish between the model-free (H_1) and model-based (H_2) strategy, we tested whether there was any transient change in decision policy immediately after the switch between the two prior-cost pairs. Unlike RSG, participants' behavior did not exhibit any transient deviation from the optimal policy after the switch (Figure 7a, Supplementary figure 7), and was better explained by the optimal policy than a purely prior-sensitive (H_3) or cost-sensitive (H_4) strategy, even with the first 25 trials after the switch (Figure 7b,c; Supplementary figure 7). Together, these results suggest that, in VMR, participants rely on a model-free optimal policy (H_1) without explicit learning of the prior and cost function. More generally, this results invites caution against interpreting Bayes-optimal behavior as evidence for knowledge about the underlying priors and the cost functions.

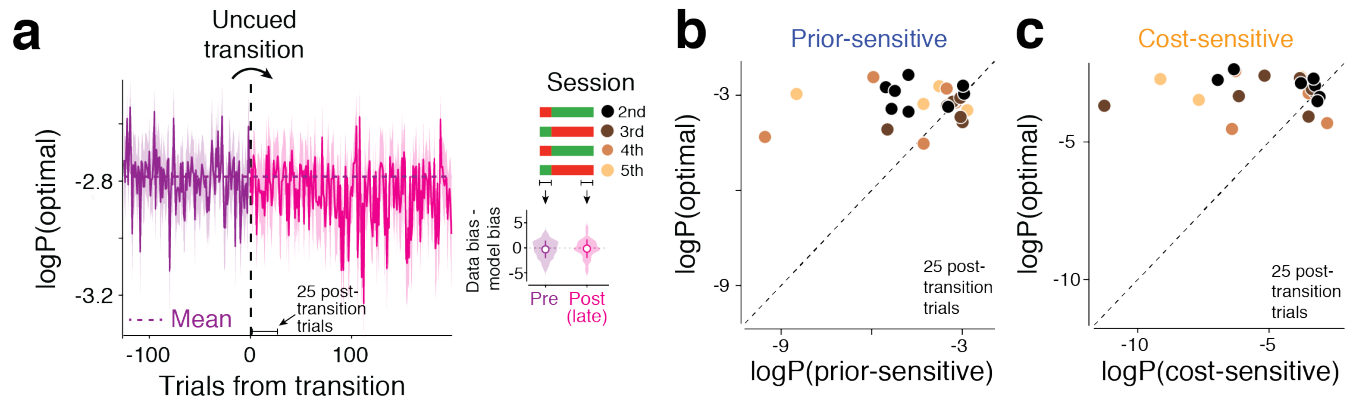


Figure 7. Behavior immediately after the switch to a new prior-cost condition in the VMR task. Results are shown in the same format as in Figure 3. Unlike the RSG task, behavior does not exhibit any transient deviation from optimal policy after the switch.

Discussion

There is a strong consensus that human decision making under uncertainty is consistent with Bayesian decision theory (BDT)^{1,3,4,22,25}. This finding has been taken as evidence that humans have internal models for the prior probability of environmental states, the likelihood of those states based on sensory evidence, and the costs (or rewards) associated with making decisions about those states. However, the success of fitting Bayesian models to behavior does not necessarily mean that the human brain establishes internal models for priors, likelihoods, and cost functions. Indeed, numerous researchers have contested the formulation of optimal behavior in terms of model-based BDT, arguing instead that the brain relies on heuristics^{27,47} and gradual learning of model-free optimal policies^{29,35,36}.

There have been numerous attempts to experimentally test the key model-based prediction of BDT. For example, some studies have changed one element of the BDT during experiments and studied how human observers adapt to the change^{41,48–50}. However, this approach inevitably changes the corresponding optimal decision policy and therefore could not be used to distinguish between model-based and model-free hypotheses. One potential approach for tackling this problem is to ask whether knowledge about the prior and cost would transfer to a new behavioral setting^{32,33}. For example, one can expose participants to two prior-cost conditions, P_1, C_1 , and P_2, C_2 , and ask whether behavior remains optimal when participants are tested in cross conditions; i.e., P_1, C_2 and P_2, C_1 . If task performance stays at the optimal level under the new pair, it indicates that participants can flexibly integrate learned priors and cost functions in accordance with BDT. However, implementation of this approach can be challenging since it requires participants to fully learn and flexibly access distinct priors and cost functions. As such, deviations from optimality using this approach may be due to bounded computational capacity that humans have in switching between internal models.

To fill this gap, we developed a novel experimental paradigm that capitalizes on the issue of model identifiability in BDT⁵¹. Since there is no one-to-one correspondence between an optimal decision policy and its underlying BDT elements (likelihood, prior, and cost function), it is possible to design Bayesian metamers that involve different prior-cost pairs but are associated with the same optimal policy. This approach parallels fruitful uses of metamers in perception^{37–39} and machine learning⁵². The key idea is that an observer whose decisions rely on a model-free optimal policy should ‘see’ the pairs as metamers because they correspond to the same policy. In contrast, an observer that makes direct use of priors and cost functions will see the two pairs as distinct and will therefore be sensitive to transitions between them.

We demonstrated the utility of our approach in the RSG and VMR tasks that humans perform nearly optimally. Switching between metamers led to transient deviations from the optimal policy in the RSG but not VMR task. These findings are consistent with the interpretation that the optimal policy in the RSG – not VMR – task relies on internal models for the prior and cost function. What factors might underlie the difference between the two tasks? One factor that differs between the two tasks is the complexity of the decision policies. Recall that the optimal policy in the VMR task was a linear mapping between measurements and estimates. It is conceivable that the sensorimotor system has powerful machinery to rapidly adapt to new linear sensorimotor mappings without the need to relearn internal models of the prior and cost function. In contrast, the nonlinear optimal policy in the RSG task may necessitate the re-learning of the prior and cost function for making adjustments to the policy. It may also be that the difference is due to inherent differences between sensorimotor timing (RSG) and sensorimotor reaching

(VMR). More experiments involving different forms of priors, cost functions, and optimal policies are needed to understand the conditions under which the brain adopts model-based versus model-free strategies^{53,54}.

An additional insight from our study of learning under metamers was the importance of learning dynamics. We found that when the learning rates for the prior and cost function are comparable, observers can learn new internal models without sacrificing performance during the learning. As a corollary, parallel learning may cause model-based learning to appear as model-free. Indeed, this observation reveals the Achilles' heel of our paradigm: one would not be able to reject the model-based approach if participants' learning rates for priors and cost functions are comparable. Therefore, our paradigm provides sufficient – not necessary – evidence for the model-based strategy.

Our framework can be flexibly extended in different ways. For example, one can further validate BDT using likelihood-cost, likelihood-prior metamers. The use of likelihood-cost metamers may be particularly valuable when the likelihood and cost can be manipulated^{6,15,42} but the prior is thought to be hard-wired, perhaps as a result of development and/or evolution¹⁰. For the prior-cost metamers that we focused on, there is a great deal of flexibility in choosing the parametric form of the prior and cost function. This flexibility can be used to validate conclusions that previous studies made about the role of a specific form of prior distribution⁵⁵ and/or cost function^{40,56} in human sensorimotor behavior. Finally, the specific design considerations for the metamer is not limited to linear shifts of the prior and/or cost function and can be adjusted based on experimental needs.

Care must be taken in interpreting the results of metameric experiments such as ours. For example, one implicit assumption of our methodology is that the model-free strategies do not involve updating the decision policy. This is a reasonable assumption if stimulus statistics associated with the original and metamer priors are sufficiently similar. However, if the stimuli associated with the new prior are widely different from the original prior, participants may have to engage in a learning process to establish the optimal decision policy for the new stimuli. This learning may present itself as a transient deviation from optimality and can thus be misinterpreted for model-based learning. This concern is unlikely to have impacted our interpretation since we made the shifts in the prior and cost function relatively small. Moreover, we found that participants' behavior deviated from optimality even when we focused our analysis only on those stimulus conditions that were present in both the original and metamer conditions (Supplementary Figure 3). However, not addressing this issue could lead to a misinterpretation of a model-free policy as model-based BDT. One strategy that can further reduce the chance of such misinterpretation is to design metamers that involve fully overlapping ranges of stimuli (i.e., have the same domain) but with different probability profiles (e.g., different variance or skewness). This strategy would ensure that participants have observed all possible stimuli in the metamer before the transition.

It is also important to carefully titrate the cost function. If the transition to the metamer leads to an overall increase or decrease in reward, participants might notice the change and initiate exploratory behavior to update the decision policy. Again, this concern is unlikely to have impacted our results since we adjusted the metamer cost function to ensure the overall expected reward was unchanged after the transition. Nevertheless, it is important to design the cost function metamers carefully so as to avoid misattributing learning of a new model-free policy to model-based learning.

Our analysis of learning dynamics in RSG suggests that updating priors and cost functions likely involve different neural systems. This conclusion is consistent with the underlying neurobiology⁵⁷. Humans and animals update their prior beliefs when observed stimuli deviate from predictions, which can be quantified in terms of a sensory prediction error (SPE)^{22,58–60}. The cerebellum is thought to play a particularly important role in supervised SPE-dependent learning of stimulus statistics in sensorimotor behaviors^{61–64}. In contrast, learning cost functions is thought to depend on computing reward prediction errors (RPE) between actual and expected reward^{65–67}. This type of learning is thought to involve the midbrain dopaminergic system⁶⁸ in conjunction with the cortico-basal ganglia circuits^{57,69–71}. Finally, the information about the prior and cost function has to be integrated to drive optimal behavior. Currently, the neural circuits and mechanisms that are responsible for this integration are not well understood^{72–74}. Future work could take advantage of our methodology to systematically probe behavioral settings that rely on model-based BDT as a rational starting point for making inquiries about the underlying neural mechanisms.

Methods

Eleven participants (age: 18-65 years, 6 male and 5 female) participated in the experiments after giving informed consent. All participants were naive to the purpose of the study, had normal or corrected-to-normal vision, and were paid for their participation. All experiments were approved by the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology.

Procedures

All 11 participants completed five experimental sessions of the Ready-Set-Go time interval reproduction task (RSG), and 10 of them completed five sessions of the visuomotor rotation task (VMR). The testing sequence for the two tasks was counterbalanced across participants. In each session, a participant was seated in a dark quiet room and asked to perform the task of interest for approximately 60 min. For both tasks, stimuli and behavioral contingencies were controlled by an open-source software (MWorks; <http://mworks-project.org/>) running on an Apple Macintosh platform.

Before the first session, participants received detailed instruction about task contingencies, and completed dozens of practice trials. We used the data from the first session to make baseline measurements of various participant-specific parameters needed for adjusting experimental parameters in the remaining sessions (see below for details). The data from the remaining sessions were used to ask whether humans use model-free versus model-based Bayesian integration strategy (Figure 1).

RSG task

Stimuli were presented on a fronto-parallel 23-inch display (distance: approximately 67 cm; refresh rate: 60 Hz; resolution: 1920 by 1200) and behavioral responses were registered using a standard Apple keyboard. Participants' task was to measure a sample time interval (t_s) between two visual flashes ('Ready' and 'Set') and subsequently initiate a delayed motor response ('Go') such that the produced interval relative to Set (t_p) would match t_s as accurately as possible. Each trial began with the presentation of a central fixation point (FP; white circle, diameter: 0.5 degree in visual angle) along with a peripheral stimulus (white circle, diameter: 0.25 deg) that was presented 5 deg to the left of FP. We asked participants to maintain their gaze on FP throughout the trial but their eye positions were not monitored. The peripheral stimulus only served as a spatial reference and was otherwise irrelevant. After a random delay (exponentially distributed; mean of 500 ms with a minimum of 250 ms), Ready and Set flashes (white circle, diameter: 0.25 deg; flash duration: 50 ms) were presented in sequence to the right and top of FP. On each trial, t_s was sampled from the prior probability distribution (see below for details about the prior). After Set, participants had to produce a matching interval by pressing the spacebar on a keyboard, and received a graded numerical feedback ('reward') below FP based on the magnitude of the error (see below for details about the cost/reward function). Trials were separated by a fixed 500 ms inter-trial interval.

Prior-cost metamers for RSG

On each trial, t_s was sampled from a prior distribution, and the numerical feedback was computed based on an experimentally imposed cost function. We measured performance in the context of two prior-cost pairs, the 'original' pair, which was introduced in the first session, and its 'metamer' that was additionally introduced on the remaining session.

We denote the original prior by $p_o(t_s)$ and the original cost function by $C_o(t_e, t_s)$. For all participants, $p_o(t_s)$ was a Gaussian distribution with a mean of 750 ms and a standard deviation of 144 ms (Figure 2c). $C_o(t_e, t_s)$ was a concave quadratic function of error ($t_e - t_s$) with a y-intercept of 100 such that participants received a maximum score of 100 when $t_e = t_s$ (Figure 2c). For errors larger than 1000 ms associated with negative scores, we truncated the cost function and set the score to 0. Mathematically, $p_o(t_s)$ and $C_o(t_e, t_s)$ were defined as follows:

$$p_o(t_s) = \frac{1}{144\sqrt{2\pi}} e^{-\frac{(t_s-750)^2}{2(144)^2}} \quad (\text{Equation 1})$$

$$C_o(t_e, t_s) = \begin{cases} 100 - \frac{9.9649}{10^5}(t_e - t_s)^2 & |t_e - t_s| < 1000^{ms} \\ 0 & \text{otherwise.} \end{cases} \quad (\text{Equation 2})$$

Following previous work⁴⁰, we used a generative model to compute the ideal observer policy associated with the original prior-cost pair. We assumed that the ideal observer measurement of t_s , denoted by t_m , is perturbed by zero-mean Gaussian noise. Based on the scalar property of timing variability⁷⁵, we additionally assumed that the standard deviation of noise scales with t_s with a constant of proportionality w_m , which we refer to as the Weber fraction for measurement. Accordingly, we formulated the likelihood function, $p(t_m|t_s)$ as $N(t_s, w_m t_s)$. We computed the ideal observer policy for the original pair, denoted by f_o^{ideal} , by integrating $p(t_m|t_s)$ with $p_o(t_s)$ and the original cost function by $C_o(t_e, t_s)$ and maximizing the expected reward:

$$f_o^{ideal}(t_m) = \arg \max_{t_e} \int p(t_m|t_s) p_o(t_s) C_o(t_e, t_s) dt_s \quad (\text{Equation 3})$$

To compare the behavior of the participants to that of the ideal observer policy, we augmented the model to include production noise such that the produced interval, t_p , was a sample from a Gaussian distribution with mean t_e , and standard deviation $w_p t_e$ ($N(t_e, w_p t_e)$) where w_p reflects additional scalar variability associated with the production⁷⁶. We fit the ideal observer model to data from the first session to compare participants' behavior to the optimal policy and to estimate each participant's w_m and w_p , which remained relatively stable throughout the entire experiment (Supplementary figure 8). Note that the scores participants received during the experiment were based on the error in t_p , not t_e , since t_e is not observable. However, using simulations, we verified that the expected cost function with t_e is similar to the cost function with t_p when the production noise is symmetric with respect to t_e (not shown). In computing the ideal observer policy, we did not take into account the fact that the quadratic cost function was truncated as errors were rarely within the truncated sidebands (mean \pm SD: 0.0064 \pm 0.0211% across subjects).

For the subsequent experimental sessions, we had to design a metamer for the original prior-cost pair. To do so, we considered a new prior, $p_m(t_s)$ and a new cost function, $C_m(t_e, t_s)$ such that the ideal observer policy for the metamer, f_m^{ideal} would be identical to f_o^{ideal} . Mathematically, this imposed the following equality constraint:

$$\arg \max_{t_e} \int p(t_m|t_s)p_o(t_s)C_o(t_e, t_s)dt_s = \arg \max_{t_e} \int p(t_m|t_s)p_m(t_s)C_m(t_e, t_s)dt_s \quad (\text{Equation 4})$$

Evidently, there are an infinitude of solutions to this problem since we can freely adjust both $p_m(t_s)$ and $C_m(t_e, t_s)$. For the purpose of our experiment, we wanted the original and its metamer to be generally similar but different enough to provide statistical power for detecting a transient learning effect. We focused on a specific form of manipulation that involved a shift in the cost function toward lower values, and commensurate adjustments to the prior to satisfy the constraint of generating a metamer. Accordingly, we defined $C_m(t_e, t_s)$ to be the same as $C_o(t_e, t_s)$ but shifted by an interval, Δt . Note that we favored this choice because of its computational simplicity but other manipulations to the cost function (e.g., introducing asymmetry) are also permissible.

Another factor we had to consider beyond matching the general form of the ideal observer policy was to make sure that the average score a participant received (the integral term in Equation 4) was similar between the original pair and its metamer. This was important because a sudden change in average reward (score) could signal a change in the environment and motivate learning a new policy. To address this point, we scaled $C_m(t_e, t_s)$ by a factor k whose value was adjusted so that the expected reward for the two pairs were the same. With these considerations in mind, the new cost function was formulated as follows:

$$C_m(t_e, t_s) = \begin{cases} k(100 - \frac{9.9649}{10^5}(t_e - t_s - \Delta t)^2) & |t_e - t_s - \Delta t| < 1000^{ms} \\ 0 & \text{otherwise.} \end{cases} \quad (\text{Equation 5})$$

Using simulations (not shown), we determined that, depending on the Weber fraction, a shift in the cost function in the order of 100 to 200 ms (larger shifts for larger Weber fractions) was sufficient for detecting learning effects. Accordingly, we set Δt to 150 ms if the fitted w_m was less than 0.2, and 200 ms otherwise. Matching the average score also provided a participant-specific value for k (mean across participants: 0.73, SD: 0.06).

Having fixed the choice of $C_m(t_e, t_s)$, we next had to design $p_m(t_s)$. To do so, we modeled $p_m(t_s)$ as a Gaussian mixture model (GMM), $\sum w_k N(\mu_k, \sigma_k)$, and estimated w_k and σ_k such that the new cost-prior pair was associated with the same ideal observer policy as the original pair (Equation 4; 'fmincon' function in MATLAB). To make this optimization process more tractable, we fixed the number of components in the GMM to 20 and fixed their corresponding mean to be equidistant between 350 to 1300 ms. Note that $p_m(t_s)$ had to be customized for each participant separately because the exact shape $p_m(t_s)$ depends on w_m .

We then used the new prior-cost pair in the subsequent sessions. For the second and fourth sessions, we started the experiment with the original pair and then switched covertly to the metamer pair. For the third and fifth sessions, the order of the original and its metamer were switched. In all these sessions, the first prior-cost pair was present for the first 170 trials, and the other pair was present for the remaining trials (approximately 590 trials). More trials were dedicated to the post-transition so that we could observe the learning-related effects for longer periods.

VMR task

Participants viewed the stimuli from above on a 21.5-inch Samsung SyncMaster SA200 display (distance: approximately 60 cm; refresh rate: 60 Hz; resolution: 1920 by 1080), and responded by controlling a custom manipulandum built from a stylus for a pen digitizer tablet (Wacom Intuous5 touch) with their hand occluded by the display. Participants' task was to use a manipulandum to move a partially occluded cursor from its initial position at the center of the display to a target positioned over the circumference of a ring around the initial position. Each trial began with the presentation of a central point (FP; red square, size: 1 deg). To proceed, participants had to use the manipulandum to move the cursor, which at this stage of the trial was visible, to the center of the screen (over FP). As soon as the cursor entered an electronic circular window (diameter: 0.5 deg) around FP, a visual target (gray circle, contrast: 40%, diameter: 1 deg) was presented briefly (duration: 500 ms) on a circular ring (radius: 10 deg, thickness: 0.1 deg, contrast: 50%) around FP. The ring remained visible throughout the trial. The position of the target was chosen from a uniform distribution between 80 to 100 deg with its midpoint aligned to 12 o'clock over the circle. Afterwards, participants had to move the cursor to the remembered location of the flashed target. The cursor was made invisible as soon as it passed the electronic window, and remained invisible throughout its movement except for a brief reappearance (gray circle, contrast: 40%, diameter: 0.5 deg, duration: 100 ms) when it was 3 deg from FP. Participants were asked to move the cursor smoothly with no interruption.

Critically, while the cursor was invisible and before its reappearance at 3 deg from FP, we applied an angular rotation (x_s) to the cursor relative to the hand position. On each trial, x_s was taken from a prior distribution (see below for details about the prior). Participants had to use their knowledge about the prior distribution and the visual feedback along the path to infer x_s , and maneuver the cursor to produce a counter rotation (x_p) that matches x_s as accurately as possible. We defined x_p as the angular distance between the target and hand position over the ring. Similar to RSG, participants received a graded numerical feedback below FP based on the magnitude of the error (see below for details about the cost/reward function). Trials were separated by a fixed 500 ms inter-trial interval. For VMR, we asked participants to use their dominant hand throughout the experiment.

Prior-cost metamers for VMR

The original prior, $p_o(x_s)$ and cost function $C_o(x_e, x_s)$ were governed by the following equations:

$$p_o(x_s) = \frac{1}{7.5\sqrt{2\pi}} e^{-\frac{(x_s - (-12))^2}{2(7.5)^2}} \quad (\text{Equation 6})$$

$$C_o(x_e, x_s) = \begin{cases} 100 - 0.111(x_e - x_s)^2 & |x_e - x_s| < 30^\circ \\ 0 & \text{otherwise.} \end{cases} \quad (\text{Equation 7})$$

To compute the ideal observer policy for the VMR task, we assumed that the measured angle, x_m is subject to zero-mean Gaussian noise, and formulated the likelihood function, $p(x_m|x_s)$ as $N(x_s, \sigma_m)$. In this setting with a Gaussian prior and a Gaussian likelihood function, f_o^{ideal} is a simple linear mapping between x_e and x_m as follows:

$$x_e = \frac{7.5^2}{7.5^2 + \sigma_m^2} x_m + \frac{\sigma_m^2}{7.5^2 + \sigma_m^2} (-12) \quad (\text{Equation 8})$$

To compare the participants behavior to the ideal observer, we additionally assumed that the produced angular correction, x_p , was also subject to zero-mean Gaussian noise with standard deviation, σ_p ; i.e., $x_p \sim \mathcal{N}(x_e, \sigma_p)$. We fit the ideal observer model to data from the first session to compare participants' behavior to the optimal policy and to estimate each participant's σ_m and σ_p , which remained relatively stable throughout the entire experiment (Supplementary figure 9). Note that we defined the cost function in terms of error in x_e , and computed the scores based on the error in x_p . Similar to the RSG task, we verified that the expected cost function with x_e is similar to the cost function with x_p when the production noise is symmetric with respect to x_e (not shown).

Our procedure for devising a metamer for the original prior-cost pair was as follows: we formulated $p_m(x_s)$ as a Gaussian distribution that was shifted by 10 deg, and $C_m(x_e, x_s)$ as a quadratic function that was shifted in the opposite direction, as follows:

$$\arg \max_{x_e} \int p(x_m | x_s) p_o(x_s) C_o(x_e, x_s) dx_s = \arg \max_{x_e} \int p(x_m | x_s) p_m(x_s) C_m(x_e, x_s) dx_s \quad (\text{Equation 9})$$

Note that the design of the metamer for the VMR is significantly easier than the RSG task because, in the VMR task, changing the mean of the prior only alters the intercept of the linear policy and not its slope, which can be readily compensated by an opposite shift in the cost function. We computed the requisite shift in the cost function (Δx) analytically by matching f_o^{ideal} to f_m^{ideal} as follows:

$$C_m(x_e, x_s) = \begin{cases} k(100 - 0.111(x_e - x_s - \Delta x)^2) & |x_e - x_s - \Delta x| < 30^\circ \\ 0 & \text{otherwise.} \end{cases} \quad (\text{Equation 10})$$

where Δx can be computed as follows:

$$\Delta x = -10 \frac{\sigma_m^2}{7.5^2 + \sigma_m^2} \quad (\text{Equation 11})$$

Finally, we used the original and metamer prior-cost pairs in the subsequent sessions using the same procedure as in the RSG task.

Models and Analysis

All analyses were performed using custom code in MATLAB (Mathworks, Inc.). We first removed outlier trials for each data set across all participants and sessions. We applied different algorithms to the two tasks as their response profile was inherently different (i.e., nonlinear metronomic function for the RSG task and linear policy for the VMR). For the RSG task, we excluded trials in which the relative error, defined as $(t_p - t_s)/t_s$, deviated more than 3 standard deviations from its mean (mean: 0.50%; SD: 0.28% across subjects). For the VMR task, we first fitted a linear regression model relating x_p and x_s , and excluded trials for which the error from the linear fit was more than 3.5 times larger than the median

absolute deviation (mean: 3.4%; SD: 2% across data sets). We verified that outlier trials were not concentrated immediately after the switch between the prior-cost pairs.

Decision policy for the model-free (H_1) and model-based (H_2) optimal strategies in the RSG task

For the original prior-cost pair, which uses a quadratic cost function, the optimal policy is to choose the mean of the posterior:

$$t_e = f_o(t_m) = \frac{\int t_s p(t_m | t_s) p_o(t_s) dt_s}{\int p(t_m | t_s) p_o(t_s) dt_s} \quad (\text{Equation 12})$$

This function captures the asymptotic optimal policy for both the model-free and model-based strategies because the two pairs were designed to be metamers. The same optimal policy can also be written in terms of the $p_m(t_s)$ as follows:

$$f_m(t_m) = \frac{\int t_s p(t_m | t_s) p_m(t_s) dt_s}{\int p(t_m | t_s) p_m(t_s) dt_s} + \Delta t \quad (\text{Equation 13})$$

Note that the shift of the cost function, Δt , adds a constant offset to the policy, which is compensated by the change in the prior.

Decision policy for prior-sensitive (H_3) and cost-sensitive (H_4) suboptimal strategies

Under H_3 , the observer only learns the new prior and assumes that there has been no change in the quadratic cost function. The corresponding decision policy is therefore to use the mean of the posterior under the new prior:

$$f_{ps}(t_m) = \frac{\int t_s p(t_m | t_s) p_m(t_s) dt_s}{\int p(t_m | t_s) p_m(t_s) dt_s} \quad (\text{Equation 14})$$

Under H_4 , the observer only learns the new cost function and assumes that there has been no change in the Gaussian prior. The corresponding decision policy can be computed as follows:

$$f_{cs}(t_m) = \frac{\int t_s p(t_m | t_s) p_o(t_s) dt_s}{\int p(t_m | t_s) p_o(t_s) dt_s} + \Delta t \quad (\text{Equation 15})$$

The policy for the prior-sensitive and cost-sensitive models has the same form as $f_m(t_m)$ and $f_o(t_m)$, respectively, but their intercepts are different from the optimal policies (Figure 2d).

Model comparison after switch

To distinguish between H_1 , H_2 , H_3 , and H_4 , we compared the log-likelihood of the data under the four models (Equations 12-15) and asked which model better captures behavior of the participants immediately after the switch between the two prior-cost pairs. The likelihood for each trial (superscript i) under each model (j) can be computed as follows:

$$p(t_p^i | t_s^i) = \int p(t_p^i | f_j(t_m), w_p) p(t_m | t_s^i, w_m) dt_m \quad (\text{Equation 16})$$

The first and second term within the integral correspond to the noisy generative process of production and measurement, respectively (w_p and w_m for its Weber fraction). $f_j(t_m)$ can be replaced by f_o (H₁/H₂: model-free and model-based optimal policy), f_{ps} (H₃: prior-sensitive), and f_{cs} (H₄: cost-sensitive) to test the data under the corresponding hypotheses.

Decision policies and model comparisons in the VMR task

Equations 12-16 also apply to the VMR task by substituting t_s , t_m , t_e , t_p with x_s , x_m , x_e , x_p , respectively.

Fitting subjective prior and cost function for re-learning strategy

To better characterize the relearning process after the prior-cost switch, we fitted the behavior after switch using an ideal observer model in which the prior and cost function could change dynamically across trials. This exercise allowed us to examine the time course of relearning and provided a basis for later analyses where we compared the learning rates between the prior and cost function. To do so, we assumed participants used a Gaussian prior and a quadratic cost function, and treated the mean of the prior, μ , and shift in the cost function, Δt , as free parameters of the model (in addition to the variance of the prior and the two Weber fractions for measurement and production; Supplementary figure 4) that were fitted based on the behavioral data after the switch. The two Weber fractions as additional parameters did not change the result in Figure 4 as a simpler model without free w_m and w_p led to qualitatively similar results (not shown). We estimated the model parameters by maximizing the likelihood of data t_p and t_s (Equation 16 with the decision policy $f_j(t_m)$ from Equation 13; 'fminsearch' function in MATLAB) under the model. We used a running window of 100 trials, which led to 1 pre-transition and 4 post-transition estimates of μ and Δt . We did not fit a subjective model to the VMR data as we could not reject the model-free strategy (Figure 7a).

Finally, we compared different models (model-free, model-based, prior-sensitive, cost-sensitive) across all data sets using Bayesian Information Criteria ($BIC = -2 \times L + N_{param} \times \log(N_{obs})$) where L is the likelihood of data given model parameters, N_{param} is the number of free parameters in the model, N_{obs} is the number of data points, i.e., trials). For model-free, prior-sensitive, and cost-sensitive models, we used the first session's w_m and w_p estimates, which remained relatively stable across sessions (Supplementary figure 8). Therefore, N_{param} was zero for all models except the subjective re-learning model (5 per each data set, 100-trial window).

Analysis of learning rates and simulation

The subjective re-learning model allowed us to capture how the internal prior and cost function is updated after the transition. We fitted an exponential function to the time course of the subjective μ and Δt to quantify and compare the corresponding learning rates for each participant. To make the fits more robust, we constrained the initial and final values of μ and Δt to match the experimentally imposed values for the pre- and post-transition (μ_{pre} , Δt_{pre} and μ_{post} , Δt_{post} , respectively). With this consideration, we only needed to fit the parameter associated with the learning rate (τ_μ , $\tau_{\Delta t}$; Figure 5a), as follows.

$$\mu^i = (\mu_{pre} - \mu_{post}) \exp\left(\frac{-i}{\tau_\mu}\right) + \mu_{post} \quad (\text{Equation 17})$$

$$\Delta t^i = (\Delta t_{pre} - \Delta t_{post}) \exp\left(\frac{-i}{\tau_{\Delta t}}\right) + \Delta t_{post} \quad (\text{Equation 18})$$

Where μ^i , Δt^i are the prior mean and cost shift in trial i after the transition. We estimated τ_μ , $\tau_{\Delta t}$ by maximizing the likelihood of data t_p and t_s (Equation 16 with the decision policy $f_j(t_m)$ from Equation 13; 'fminsearch' function in MATLAB). Similar to the subjective model fitting, we allowed both the mean and variance of the prior to change throughout learning.

We also performed simulations to examine how the learning rates for μ and Δt influence performance (Figure 5f). The procedure for each simulation was as follows: 1) we chose specific values for the learning rates, τ_μ and $\tau_{\Delta t}$; 2) we computed μ and Δt as a function of time governed by those learning rates; 3) we computed the moment by moment value of expected reward (EC) as a function of the instantaneous values of μ and Δt as follows:

$$EC^i = \int_{t_p} \int_{t_m} C(t_p, t_s^i) p(t_p | f_j(t_m), w_p) p(t_m | t_s^i, w_m) dt_m dt_p \quad (\text{Equation 19})$$

In this equation, we computed EC^i (expected cost in trial i) in the context of the metamer prior-cost pair for an observer with $w_m = 0.14$ and $w_p = 0.08$, close to averages across subjects (Supplementary Figure 8). In this equation, different values of μ and Δt influence EC through the decision policy $f_j(t_m)$ (Equation 13). We also measured the overall average reward (marginalizing over time) for different choices of τ_μ and $\tau_{\Delta t}$.

References

- 1 Bulthoff HH. Bayesian decision theory and psychophysics. *Perception as Bayesian inference* 1996; **123**.https://books.google.com/books?hl=en&lr=&id=_cTLCgAAQBAJ&oi=fnd&pg=PA123&dq=yuille+bulthoff&ots=CfjLzBrBUU&sig=sHmt5A4SmdVibXq5AG5QVlbCiJ0.
- 2 Berger JO. *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media, 2013.
- 3 Körding KP, Wolpert DM. Bayesian decision theory in sensorimotor control. *Trends Cogn Sci* 2006; **10**: 319–326.
- 4 Dijkstra TMH, de Vries B, Heskes TM, Zoeter OR. A Bayesian decision-theoretic framework for psychophysics. International Society for Bayesian Analysis, 2006.
- 5 Landy MS, Maloney LT, Johnston EB, Young M. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res* 1995; **35**: 389–412.
- 6 Stocker AA, Simoncelli EP. Noise characteristics and prior expectations in human visual speed perception. *Nat Neurosci* 2006; **9**: 578–585.
- 7 Girshick AR, Landy MS, Simoncelli EP. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat Neurosci* 2011; **14**: 926–932.
- 8 Maloney LT, Zhang H. Decision-theoretic models of visual perception and action. *Vision Res* 2010; **50**: 2362–2374.
- 9 Knill DC. Mixture models and the probabilistic structure of depth cues. *Vision Res* 2003; **43**: 831–854.
- 10 Adams WJ, Graf EW, Ernst MO. Experience can change the ‘light-from-above’ prior. *Nat Neurosci* 2004; **7**: 1057–1058.
- 11 Körding KP, Wolpert DM. Bayesian integration in sensorimotor learning. *Nature* 2004; **427**: 244–247.
- 12 Trommershäuser J, Maloney LT, Landy MS. Decision making, movement planning and statistical decision theory. *Trends Cogn Sci* 2008; **12**: 291–297.
- 13 Najemnik J, Geisler WS. Optimal eye movement strategies in visual search. *Nature* 2005; **434**: 387–391.
- 14 Verstynen T, Sabes PN. How each movement changes the next: an experimental and theoretical study of fast adaptive priors in reaching. *J Neurosci* 2011; **31**: 10050–10059.
- 15 Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 2002; **415**: 429–433.
- 16 Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 2004; **14**: 257–262.
- 17 Angelaki DE, Gu Y, DeAngelis GC. Multisensory integration: psychophysics,

neurophysiology, and computation. *Curr Opin Neurobiol* 2009; **19**: 452–458.

18 Hillis JM, Ernst MO, Banks MS, Landy MS. Combining sensory information: mandatory fusion within, but not between, senses. *Science* 2002; **298**: 1627–1630.

19 Chater N, Tenenbaum JB, Yuille A. Probabilistic models of cognition: where next? *Trends Cogn Sci* 2006; **10**: 292–293.

20 Tenenbaum JB, Griffiths TL, Kemp C. Theory-based Bayesian models of inductive learning and reasoning. *Trends Cogn Sci* 2006; **10**: 309–318.

21 Griffiths TL, Tenenbaum JB. Optimal predictions in everyday cognition. *Psychol Sci* 2006; **17**: 767–773.

22 Doya K, Ishii S, Pouget A, Rao RPN. *Bayesian Brain: Probabilistic Approaches to Neural Coding*. MIT Press, 2007.

23 Rao RPN. Bayesian computation in recurrent neural circuits. *Neural Comput* 2004; **16**: 1–38.

24 Pouget A, Beck JM, Ma WJ, Latham PE. Probabilistic brains: knowns and unknowns. *Nat Neurosci* 2013; **16**: 1170–1178.

25 Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 2004; **27**: 712–719.

26 Friston K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 2010; **11**: 127–138.

27 Bowers JS, Davis CJ. Bayesian just-so stories in psychology and neuroscience. *Psychol Bull* 2012; **138**: 389–414.

28 Marcus GF, Davis E. How robust are probabilistic models of higher-level cognition? *Psychol Sci* 2013; **24**: 2351–2360.

29 Jones M, Love BC. Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behav Brain Sci* 2011; **34**: 169–188.

30 Laquittaine S, Gardner JL. A Switching Observer for Human Perceptual Estimation. *Neuron* 2018; **97**: 462–474.e6.

31 Raphan M, Simoncelli EP, Scholkopf B, Platt J, Hoffman T. Learning to be Bayesian without Supervision. In: *Neural Information Processing Systems*. 2006, pp 1145–1152.

32 Ma WJ, Jazayeri M. Neural coding of uncertainty and probability. *Annu Rev Neurosci* 2014; **37**: 205–220.

33 Maloney LT, Mamassian P. Bayesian decision theory as a model of human visual perception: testing Bayesian transfer. *Vis Neurosci* 2009; **26**: 147–155.

34 Simoncelli EP. Optimal estimation in sensory systems. *The Cognitive Neurosciences, IV* 2009; : 525–535.

- 35 Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. MIT Press, 2018.
- 36 Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. Building machines that learn and think like people. *Behav Brain Sci* 2017; **40**: e253.
- 37 Evans RM. *The perception of color*. John Wiley & Sons, 1974.
- 38 Freeman J, Simoncelli EP. Metamers of the ventral stream. *Nat Neurosci* 2011; **14**: 1195–1201.
- 39 Norman-Haignere SV, McDermott JH. Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLoS Biol* 2018; **16**: e2005127.
- 40 Jazayeri M, Shadlen MN. Temporal context calibrates interval timing. *Nat Neurosci* 2010; **13**: 1020–1026.
- 41 Miyazaki M, Nozaki D, Nakajima Y. Testing Bayesian models of human coincidence timing. 2005; **94**: 395–399.
- 42 Körding KP, Wolpert DM. Bayesian integration in sensorimotor learning. *Nature* 2004; **427**: 244–247.
- 43 Remington ED, Parks TV, Jazayeri M. Late Bayesian inference in mental transformations. *Nat Commun* 2018; **9**: 4419.
- 44 Egger SW, Jazayeri M. A nonlinear updating algorithm captures suboptimal inference in the presence of signal-dependent noise. *Sci Rep* 2018; **8**: 12597.
- 45 Cicchini GM, Arrighi R, Cecchetti L, Giusti M, Burr DC. Optimal encoding of interval timing in expert percussionists. *J Neurosci* 2012; **32**: 1056–1060.
- 46 Acerbi L, Wolpert DM, Vijayakumar S. Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS Comput Biol* 2012; **8**: e1002771.
- 47 Gigerenzer G, Brighton H. Homo heuristics: why biased minds make better inferences. *Top Cogn Sci* 2009; **1**: 107–143.
- 48 Berniker M, Voss M, Kording K. Learning priors for Bayesian computations in the nervous system. *PLoS One* 2010; **5**. doi:10.1371/journal.pone.0012686.
- 49 Hudson TE, Maloney LT, Landy MS. Optimal compensation for temporal uncertainty in movement planning. *PLoS Comput Biol* 2008; **4**: e1000130.
- 50 Sohn H, Lee S-H. Dichotomy in perceptual learning of interval timing: calibration of mean accuracy and precision differ in specificity and time course. *J Neurophysiol* 2013; **109**: 344–362.
- 51 Acerbi L, Ma WJ, Vijayakumar S. A Framework for Testing Identifiability of Bayesian Models of Perception. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ (eds). *Advances in Neural Information Processing Systems 27*. Curran Associates, Inc., 2014, pp 1026–1034.
- 52 Feather J, Durango A, Gonzalez R, McDermott J. Metamers of neural networks reveal

divergence from human perceptual systems. In: Wallach H, Laroche H, Beygelzimer A, Alché-Buc F, Fox E, Garnett R (eds). *Advances in Neural Information Processing Systems* 32. Curran Associates, Inc., 2019, pp 10078–10089.

53 Kool W, Cushman FA, Gershman SJ. When Does Model-Based Control Pay Off? *PLoS Comput Biol* 2016; **12**: e1005090.

54 Kool W, Gershman SJ, Cushman FA. Cost-Benefit Arbitration Between Multiple Reinforcement-Learning Systems. *Psychol Sci* 2017; **28**: 1321–1333.

55 Zhang H, Daw ND, Maloney LT. Human representation of visuo-motor uncertainty as mixtures of orthogonal basis distributions. *Nat Neurosci* 2015; **18**: 1152–1158.

56 Körding KP, Wolpert DM. The loss function of sensorimotor learning. *Proc Natl Acad Sci U S A* 2004; **101**: 9839–9842.

57 Doya K. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Netw* 1999; **12**: 961–974.

58 Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 1999; **2**: 79–87.

59 Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. Canonical microcircuits for predictive coding. *Neuron* 2012; **76**: 695–711.

60 Izawa J, Shadmehr R. Learning from sensory and reward prediction errors during motor adaptation. *PLoS Comput Biol* 2011; **7**: e1002012.

61 Narain D, Remington ED, Zeeuw CID, Jazayeri M. A cerebellar mechanism for learning prior distributions of time intervals. *Nat Commun* 2018; **9**: 469.

62 Raymond JL, Medina JF. Computational Principles of Supervised Learning in the Cerebellum. *Annu Rev Neurosci* 2018; **41**: 233–253.

63 Herzfeld DJ, Kojima Y, Soetedjo R, Shadmehr R. Encoding of action by the Purkinje cells of the cerebellum. *Nature* 2015; **526**: 439–442.

64 Brooks JX, Carriot J, Cullen KE. Learning to expect the unexpected: rapid updating in primate cerebellum during voluntary self-motion. *Nat Neurosci* 2015; **18**: 1310–1317.

65 Watabe-Uchida M, Eshel N, Uchida N. Neural Circuitry of Reward Prediction Error. *Annu Rev Neurosci* 2017; **40**: 373–394.

66 Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, Ornelas S *et al*. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* 2019; **570**: 509–513.

67 Mohebi A, Pettibone JR, Hamid AA, Wong J-MT, Vinson LT, Patriarchi T *et al*. Dissociable dopamine dynamics for learning and motivation. *Nature* 2019; **570**: 65–70.

68 Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. *Science* 1997; **275**: 1593–1599.

- 69 Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 2005; **8**: 1704–1711.
- 70 Lee D, Seo H, Jung MW. Neural basis of reinforcement learning and decision making. *Annu Rev Neurosci* 2012; **35**: 287–308.
- 71 Athalye VR, Carmena JM, Costa RM. Neural reinforcement: re-entering and refining neural dynamics leading to desirable outcomes. *Curr Opin Neurobiol* 2020; **60**: 145–154.
- 72 Sohn H, Narain D, Meirhaeghe N, Jazayeri M. Bayesian Computation through Cortical Latent Dynamics. *Neuron* 2019; **103**: 934–947.e5.
- 73 Darlington TR, Beck JM, Lisberger SG. Neural implementation of Bayesian inference in a sensorimotor behavior. *Nat Neurosci* 2018; **21**: 1442–1451.
- 74 Akrami A, Kopec CD, Diamond ME, Brody CD. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature* 2018; **554**: 368–372.
- 75 Malapani C, Fairhurst S. Scalar Timing in Animals and Humans. *Learn Motiv* 2002; **33**: 156–176.
- 76 Jazayeri M, Shadlen MN. Temporal context calibrates interval timing. *Nat Neurosci* 2010; **13**: 1020–1026.