Synaptic computation underlying probabilistic inference

Alireza Soltani[1,2] & Xiao-Jing Wang[1]

We propose that synapses may be the workhorse of the neuronal computations that underlie probabilistic reasoning. We built a neural circuit model for probabilistic inference in which information provided by different sensory cues must be integrated and the predictive powers of individual cues about an outcome are deduced through experience. We found that bounded synapses naturally compute, through reward-dependent plasticity, the posterior probability that a choice alternative is correct given that a cue is presented. Furthermore, a decision circuit endowed with such synapses makes choices on the basis of the summed log posterior odds and performs near-optimal cue combination. The model was validated by reproducing salient observations of, and provides insights into, a monkey experiment using a categorization task. Our model thus suggests a biophysical instantiation of the Bayesian decision rule, while predicting important deviations from it similar to the 'base-rate neglect' observed in human studies when alternatives have unequal prior probabilities.

Decision making often relies on our ability to combine information from different sources and to make inferences even when the relationship between cues and outcomes is not deterministic. For instance, in the so-called weather prediction task that is commonly used in cognitive neuroscience, a categorical choice (rain or sunshine) can be predicted only probabilistically on the basis of given cues[1–5]. Such a decision is challenging not only owing to its probabilistic character but also because a single choice is preceded by many cues, so it is not obvious how to deduce correctly cue–outcome associations (for example, identifying an allergenic or poisonous substance after consuming a few food items and getting sick). Little is known about the neural computations underlying this cognitive ability of probabilistic reasoning.

A recent study suggested that monkeys are capable of some forms of probabilistic inference and revealed neural correlates of this ability at the single-cell level in the lateral intraparietal cortex (LIP)[6]. In particular, this neural activity encodes the combination of information from different cues (visual shapes) in terms of the log likelihood ratio (log LR), a quantity which the monkeys seemed to learn through experience with the cues and use to combine information in order to make a decision on each trial. This neurophysiological finding supports the theoretical proposal that log LR provides a quantity suitable for the accumulation of sensory evidence[7,8]. However, it raises the question of how such a quantity could be computed and learned biophysically.

We propose that quantities such as the likelihood or posterior probability can be learned and encoded by synapses that have bounded weights and undergo reward-dependent Hebbian plasticity[9–11]. The computational implications of bounded synapses have only begun to be recognized. In particular, previous work has shown that the

capacity of long-term memory storage depends notably on whether synapses are bounded or not[12,13]. In our model, trial-by-trial decision making is determined by statistical sampling of stochastic neural dynamics[14–19]; firing activity of single cells correlates with conditional reward probabilities because neurons are driven by bounded synapses that learn probabilistic cue-outcome associations.

We show that in a simulated probabilistic inference task, these synapses can estimate the naive posterior probability—that is, the posterior probability that a choice alternative is assigned a reward given that a cue is presented in any combination of cues. Furthermore, in a decision circuit, the choice behavior is determined by the difference in the inputs associated with each choice option, which is approximately proportional to the sum of the log posterior odds for the presented cues. The cue combination is thus near-optimal (that is, according to the Bayes rule) when the prior probabilities of reward assignment on each choice alternative are equal. However, when priors are not equal, the model predicts specific deviations that can directly be tested experimentally. Such deviations from the Bayes rule can explain the 'base-rate neglect' effect observed in human behavioral studies[20]. Overall, our model reproduces salient behavioral and single-unit neural data[6] and provides insights into the neural mechanisms of three key computational processes: inference, cue combination and probabilistic decision making.

RESULTS
Learning posteriors by plastic synapses
In this section we show how plastic synapses are able to estimate probabilistic quantities such as posteriors. We assume that individual plastic synapses are binary (with a depressed and a potentiated state); hence, the strength of a set of plastic synapses can be

[1]Department of Neurobiology and Kavli Institute of Neuroscience, Yale University School of Medicine, New Haven, Connecticut, USA. [2]Division of Biology, California Institute of Technology, Pasadena, California, USA. Correspondence should be addressed to A.S. (soltani@caltech.edu) or X.-J.W. (xjwang@yale.edu).

Figure 1 Schematic of the model and posterior computation by plastic synapses when a single cue is presented on each trial. (**a**) Schematic of the three-layer model. The first layer consists of cue-selective neural populations, each is activated upon the presentation of a cue (a shape in the weather prediction task). The sensory cue–selective neurons provide, through some synapses endowed with reward-dependent Hebbian plasticity, inputs to two neural populations in an intermediate layer that encode reward values of two choice alternatives (action values). Combination of cues is accomplished through convergence of cue-selective neurons onto action value–encoding neurons. The latter project to excitatory and inhibitory neural populations in a decision making circuit (gray box). The choice (A or B) is determined by which of the two decision neural populations wins competition on a trial. Depending on the reward schedule, a chosen action may be rewarded or not. The presence (respectively absence) of a modulatory reward signal leads to potentiation (respectively depression) of plastic synapses. (**b**) Steady state of the synaptic strength as a function of the posterior probability for different values of the learning rate ratio $r = q_+ / q_-$ (equation (1)). When $r = 1$, the synaptic strength is equal to the posterior. (**c**) Difference in the steady state of the synaptic strengths as a function of the difference in the posteriors (left panel) and of the log posterior odds (right panel), for different learning rate ratios. Dashed lines show linear fits for the values of posterior between 0.2 and 0.8. The inset shows the relationship between $\log_{10}(x / (1 - x))$ and $x - (1 - x)$ over the same range, where $x = P(A \mid S_i)$ and $P(B \mid S_i) = 1 - P(A \mid S_i) = 1 - x$.
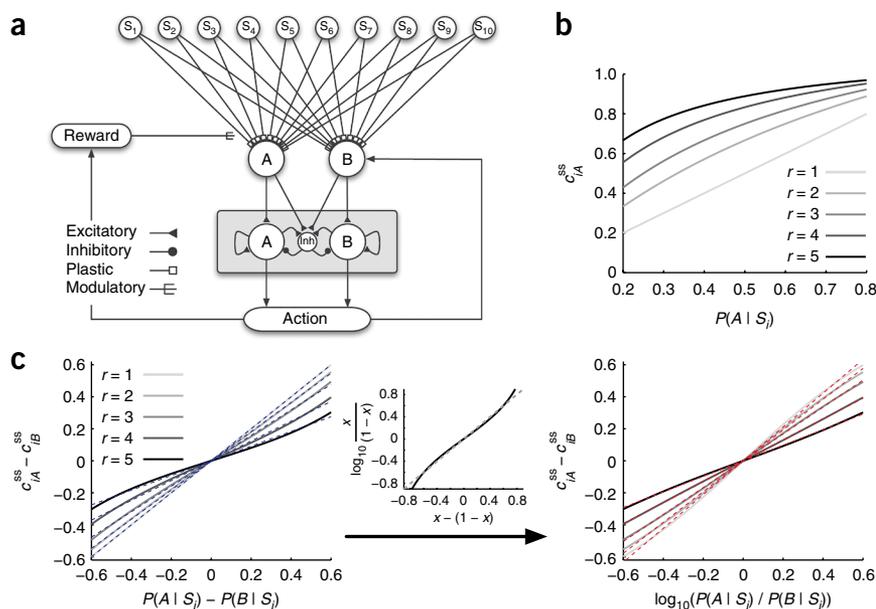
quantified by the fraction of synapses in the potentiated state[9–11,21,22]. This quantity is called the 'synaptic strength' and denoted as c_{iA} or c_{iB}, for the set of synapses from sensory neurons selective for cue S_i onto action value–encoding neurons selective for choice A or B, respectively (**Fig. 1a**).

Plastic synapses learn cue–outcome contingencies through stochastic reward-dependent Hebbian modifications[9–11] (see Online Methods for details of the learning rule). That is, at the end of the trial, only the sets of plastic synapses from sensory neurons selective for the presented cues onto action value–encoding neurons selective for the chosen action are updated, with the direction (potentiation or depression) depending on the choice outcome: if the choice of the model is rewarded, synapses in the depressed state make a transition to the potentiated state with a probability q_+; otherwise, they make a transition in the reverse direction with a probability q_-.

Consider a simple situation where only one cue (S_i) is presented on each trial, and it determines the reward probability for each of the two alternative responses, $P(A \mid S_i)$ and $P(B \mid S_i) = 1 - P(A \mid S_i)$. In this case, the reward assignment is independent of the choice selection and the probability that a set of synapses is potentiated (say, for synapses selective for cue S_i and choice A) is equal to the product of three probabilities: the probability that cue S_i is presented, $P(S_i)$; the probability that choice A is selected when cue S_i is presented, $P_A(S_i)$; and the probability that choice A is assigned a reward given cue S_i is presented, $P(A \mid S_i)$. The probability of depression for the same set of synapses is $P(S_i) \times P_A(S_i) \times (1 - P(A \mid S_i))$.

Through ongoing learning, the synaptic strength for each set of plastic synapses eventually reaches a steady-state value. If the learning rates are small, the steady state of the synaptic strength can be computed by setting the overall change equal to zero:

$$\Delta c_{iA} = q_+(1 - c_{iA}) \times P(S_i) \times P_A(S_i) \times P(A \mid S_i) - q_- c_{iA}$$
$$\times P(S_i) \times P_A(S_i) \times (1 - P(A \mid S_i)) = 0$$

which gives an expression for the steady state of the synaptic strength

$$c_{iA}^{ss} = \frac{rP(A \mid S_i)}{1 + (r - 1)P(A \mid S_i)} \tag{1}$$

where r is the learning rate ratio ($r = q_+ / q_-$). Therefore, when each cue is presented alone, the steady state is independent of the choice behavior (that is, $P_A(S_i)$).

In the special case of equal potentiation and depression rates ($r = 1$), the steady state of the synaptic strength is equal to the posterior probability, $c_{iA}^{ss} = P(A \mid S_i)$ (**Fig. 1b**). In general, when the learning rates are not equal, the synaptic strength is a nonlinear monotonic function of the posterior probability (**Fig. 1b**).

Computation of log posterior odds

In our model, the decision circuit (**Fig. 1a**) generates a categorical choice (A or B) stochastically on single trials, with a probability which is a sigmoid function of the difference in the overall inputs to its selective pools (the differential input)[9–11,14]. Because cue-selective neurons fire at a similar rate, the differential input is solely determined by the difference in the synaptic strengths from the action value–encoding neurons onto the decision neurons. Using equation (1), we can compute the difference in the synaptic strengths ($\Delta c_i^{ss} \equiv c_{iA}^{ss} - c_{iB}^{ss}$) for cue S_i

$$\Delta c_i^{ss} = \frac{r(P(A \mid S_i) - P(B \mid S_i))}{r + (r - 1)^2 P(A \mid S_i) P(B \mid S_i)} \tag{2}$$

This formula can be simplified by observing that the second term in the denominator, $k = (r - 1)^2 P(A \mid S_i) P(B \mid S_i)$, is zero when $r = 1$, and its variation is negligible (compared to the first term r) provided that r is not too large and that values of posterior probabilities are in an intermediate range (for $0.2 \leq P(A \mid S_i) \leq 0.8$, $4(r - 1)^2 / 25 \leq k \leq (r - 1)^2 / 4$). Since k is roughly constant, the difference in the steady state of the synaptic
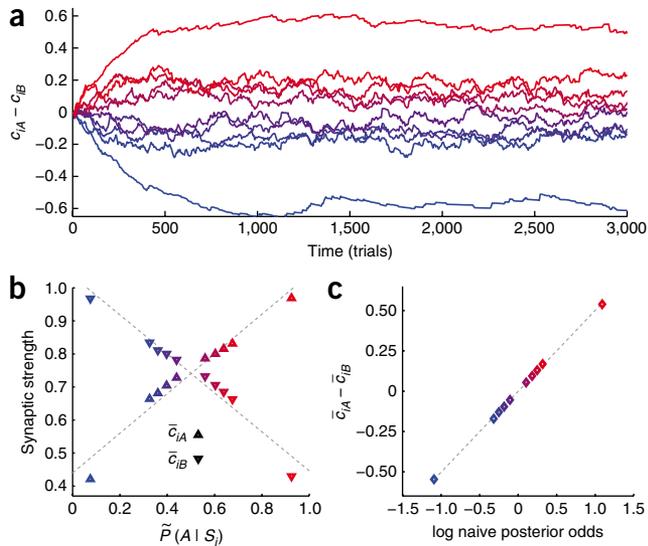
Figure 2 Posterior computation by plastic synapses when multiple cues are presented on each trial. (**a**) Time course of learning in plastic synapses. Shown is the difference in the strengths of synapses from a cue-selective neural population onto the two action value–encoding neural populations (A and B). In all figures, different color shades from blue to red correspond to shapes S_1 to S_{10}. All synaptic strengths are initially set to 0.5; they reach their steady states after a few hundred trials. (**b**) Average synaptic strength for each set of synapses plotted as a function of the naive posterior probability that alternative A is assigned a reward, given that shape S_i is presented in any pattern. (**c**) Difference in the average synaptic strengths is linearly proportional to the log naive posterior odds.

strengths is proportional to the difference in posterior probabilities for the two choice alternatives (**Fig. 1c**)

$$\Delta c_i^{ss} \approx \frac{r}{r+k}(P(A\,|\,S_i) - P(B\,|\,S_i)) \tag{3}$$

Furthermore, we note that $x - (1-x) \approx \log_{10}(x/(1-x))$ if $0.2 \le x \le 0.8$ (**Fig. 1c**). Therefore, for the intermediate range of posteriors where the model's choice behavior is stochastic, the difference in the synaptic strengths is linearly proportional to the log posterior odds (**Fig. 1c**):

$$\Delta c_i^{ss} \approx \frac{r}{r+k}\log_{10}\frac{P(A\,|\,S_i)}{P(B\,|\,S_i)} \tag{4}$$

For smaller or larger values of posteriors, the choice behavior is deterministic (the probability of choosing A is close to 0 or 1). Of particular note, equation (4) holds in the general case of unequal learning rates, when c_{iA}^{ss} is a nonlinear function of $P(A\,|\,S_i)$ (**Fig. 1b**).

In summary, when the choice outcome is based on a single cue, synapses endowed with realistic reward-dependent plasticity are capable of estimating quantities such as posteriors, and a decision network driven by such synapses can make decision according to log posterior odds.

Summation of log posterior odds

What happens if a choice outcome is preceded by several cues? To address this question, we considered a probabilistic categorization task known as the weather prediction task in which four shapes precede a selection between two (A = red, B = green) response targets on each trial[6]. These shapes are selected randomly from a set of ten distinguishable shapes (S_i, $i = 1, 2, ..., 10$), each of which is allocated a unique weight of evidence (WOE) relating to the probability of reward assignment on one of the two choice targets

$$\text{WOE} = \log_{10}\frac{P(A\,|\,S_i)}{P(B\,|\,S_i)}$$

The computer assigns a reward to one of the two alternative choices with a probability that depended on the sum of the WOEs from all the shapes presented on a given trial (see Online Methods for more details).

In model simulations of the weather prediction task, on each trial, pools of sensory neurons selected for the presented cues are activated and converge onto action value-encoding neurons (**Fig. 1a**). At the

end of the trial, all sets of plastic synapses from neurons selective for the presented shapes onto value-encoding neurons selective for the chosen action are updated independently of their role in decision making. As a result, synaptic changes for different shapes become correlated. Even though there are 10^4 stimulus patterns, we found that it took only a few hundred trials for the synaptic strengths to reach their average (steady-state) values (**Fig. 2a**), on a time scale largely set by the learning rates. This means that, after a few hundred trials, the model is able to correctly perform the task while plastic synapses continue to fluctuate (around their steady states) owing to ongoing learning.

As in equation (1), we can find an expression for the steady state of the synaptic strength (see **Supplementary Note 1**). Simulation results showed it to be approximately a linear function of the 'naive' posterior probability (**Fig. 2b**). Naive posterior probability, $\tilde{P}(A\,|\,S_i)$, is the conditional probability that alternative A is assigned a reward given that shape S_i is presented in any one epoch. It is the generalization of posterior when more than one cue precedes an outcome, assuming independence between the evidence provided by each cue. It follows from the above mathematical reasons that the difference in the synaptic strengths is approximately a linear function of the log naive posterior odds (**Fig. 2c**)

$$\Delta c_i^{ss} = \alpha \log_{10}\frac{\tilde{P}(A\,|\,S_i)}{\tilde{P}(B\,|\,S_i)} \tag{5}$$

where the linear fit yielded $\alpha = 0.48$.

Because the convergence of sensory neurons onto action value–encoding neurons naturally sums the currents through sets of plastic synapses related to presented cues, the overall differential input to decision neurons is given by the sum of log naive posterior odds. Thus, this model provides a natural mechanism for integrating evidence in terms of log posterior odds.

Simulation of weather prediction task: behavioral results

The results reported above are general, suggesting that our model endowed with the proposed reward-dependent learning rule is broadly applicable to probabilistic decision making tasks that involve inference. To test whether this model can account for behavioral performance as well as neural activity data, we simulated a monkey experiment[6].

Computer simulations followed the experimental protocol (see Online Methods for details). Because there is intrinsic noise in the neural circuit, the model's choice can vary from trial to trial even if the synaptic strengths are identical. In addition, synapses are updated at the end of each trial, hence the leverage that each presented shape has on decision making changes from trial to trial, leading to dynamic adjustment of choice behavior over time. The adaptive choice behavior of the model is described by the psychometric function (**Fig. 3a**), where the probability of selecting A is plotted against the sum of the WOEs assigned to individual shapes in a pattern. Therefore, the

Figure 3 Choice behavior of the model and the subjective weight of evidence in the weather prediction task. (**a**) Probability of choosing alternative *A* as a function of the evidence favoring this alternative for all patterns with finite WOEs. The evidence is equal to the sum of the WOE of all shapes in a pattern. For clarity, the sums of the WOE values for all patterns were binned into 0.1 intervals. Gray curve, fitted logistic function; black curves, performance of strict and probabilistic (prob.) Bayesian observers. (**b**) Subjective weight of evidence for each shape as a function of the WOE assigned to that shape.

model reproduces the main behavioral observation of ref. 6: namely, the monkeys selected each alternative stochastically based on the combined evidence provided by all presented shapes in a pattern, with a probability that is approximately a sigmoid function of the summed WOEs (see Fig. 1b in ref. 6).

The psychometric function quantifies the influence of combinations of shapes (and not the individual shapes) on the choice behavior. Following ref. 6, we used a logistic regression model (see equation (13) in Online Methods) to estimate the influence of individual shapes on the choice behavior. These regression coefficients, called the subjective weights of evidence (SWOEs), are shown in **Figure 3b**. As in the experimental findings, the SWOEs are smaller than the assigned WOEs (see Fig. 1c in ref. 6). In particular, the SWOEs for the trump shapes (that is, shapes with infinite WOE) are finite (compare **Supplementary Fig. 1** with Supplementary Fig. 2 in ref. 6). Moreover, we found that the SWOEs did not depend on the time interval (epoch) in which the shape was presented, as observed experimentally (compare **Supplementary Fig. 2** with Supplementary Fig. 3 in ref. 6).

Our results explain the following two main observations regarding the choice behavior. First, why does the model underestimate the WOEs? This happens because the SWOE of the model is proportional to the log naive posterior odds, which is less than the WOE because of the concurrence of different shapes on each trial (see **Supplementary Fig. 3**). Second, how does the model combine information from multiple cues, and why is the choice behavior approximately a sigmoid function of the summed evidence provided by shapes in a pattern (**Fig. 3a**)? For a given pattern, the choice behavior of the model is approximately a sigmoid function of the overall differential synaptic input, $\sum_i c_{iA}^{ss} - c_{iB}^{ss}$ (see **Supplementary Fig. 4** and equation (6) in **Supplementary Note 2**). Consequently, the choice behavior is a sigmoid function of the sum of the log naive posterior odds, using equation (5):

$$P_A(C^t) = \frac{1}{1 + \exp\left(-\dfrac{\alpha}{\sigma} \sum_i \log_{10} \dfrac{\tilde{P}(A \mid S_i)}{\tilde{P}(B \mid S_i)}\right)} \qquad (6)$$

where $P_A(C^t)$ is the probability of selecting *A* given pattern C^t (that is, all patterns with the sum WOE equal to *t*) is presented the sum is over all shapes in such patterns, and $1/\sigma$ quantifies the sensitivity of the choice behavior on the differential synaptic input. We found that the log naive posterior odds was linearly proportional to the WOE for the non-trump

shapes (**Supplementary Fig. 3c**). Therefore, the psychometric function shown in **Figure 3a** is a sigmoid function of the summed WOEs of all shapes (for patterns that do not contain trump shapes).

It is instructive to compare our model with an ideal Bayesian observer who can directly learn the likelihood associated with each pattern. The latter would combine the likelihood ratio associated with a given pattern and the prior odds (that each alternative is assigned a reward) to obtain the posterior odds for that pattern (see **Supplementary Note 3** for another scenario). The posterior odds then can be used to make a decision according to different decision rules; for example, strict Bayesian (selecting the alternative with the larger posterior) or probabilistic Bayesian (matching the probability of choice with the posterior). The choice behavior of these two types of Bayesian observer is shown in **Figure 3a**. We found that the reward rate (that is, the percentage harvested of the assigned rewards) for our model, probabilistic Bayesian and strict Bayesian observers to be equal to 79%, 84% and 89%, respectively. Therefore, the reward rate of our model is lower than a Bayesian observer who has direct access to the posteriors associated with each pattern without underestimating the WOEs. However, the model performance is only slightly different from a probabilistic Bayesian observer.

Neural activity correlates of probabilistic inference

To link decision making with its underlying neural activity, we examined how shape presentation influences the firing rates of decision neurons, and hence the model's choice behavior. As shown in sample neural traces (**Fig. 4**), the activity of decision neurons is driven by value-encoding neurons and is thereby influenced by the WOEs of the presented shape on each trial.

We analyzed how the neural activity and the cumulative evidence co-vary in time. Among different ways to measure evidence provided by presented shapes, we chose to perform the same analysis as in ref. 6, using the log LR that selection of red or green target is accompanied by reward after the presentation of *n* shapes (see Online Methods for details). We observed a graded dependence on the log LR (**Fig. 5a**). Moreover, the average activity in each epoch is a linear function of the average log LR in that epoch (compare with Figs. 4 and 5a in ref. 6). We also computed the incremental change in the population firing rate across successive epochs of shape presentation, and we found that this change was proportional to the average change in the log LR (Δlog LR) caused by the presentation of a new shape (see **Supplementary Fig. 5**). All of these simulation results are similar to the results observed experimentally in LIP neurons of behaving monkeys[6].

Because the neuronal activity was also strongly modulated by the choice on each trial, we performed the same analysis on the data divided into two groups depending on the choice of the model on each trial (**Fig. 5c** and **Supplementary Fig. 6**). We found that the baseline of the neural activity was higher when the choice was the preferred target. Conversely, modulation by evidence of neural activity was weaker when the choice was the preferred target. This is qualitatively similar to the observed modulations of LIP neurons when the choice was toward the response field versus away from it (see Supplementary Figs. 7 and 8 in ref. 6).

To conclude, neural activity in our model reproduced the main physiological observations from LIP in the monkey experiment[6]. These results demonstrate that empirically observed neural correlates of probabilistic quantities such as likelihoods may be interpreted in terms of synaptic rather than neuronal computations.

Model prediction: effect of prior probability

The behavioral and neural data of the weather prediction experiment were reported in terms of likelihood ratios[6]. Because in this experiment the prior reward probability was the same for the two choice
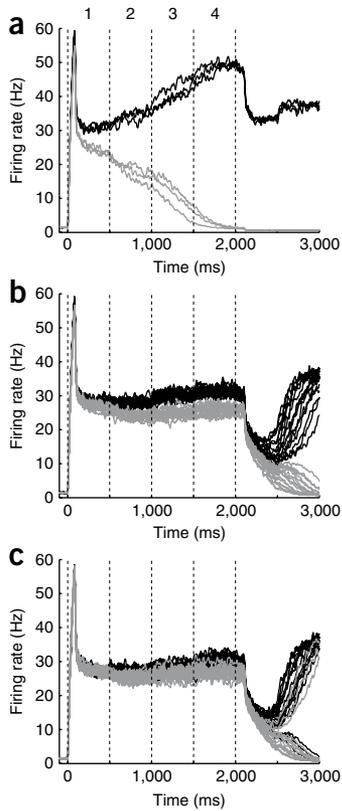
Figure 4 Model neural population activity during the weather prediction task. (**a**) Firing activity of two (black, *A*; gray, *B*) choice-selective populations in the decision-making network on a few sample trials. On these trials, the trump shape favoring *A* (with infinite WOE) is presented in epochs 1–3; in epoch 4, any one of the five shapes providing evidence in favor of *A* is presented. Time zero: onset of choice targets, fixation point and the first visual shape stimulus. Four epochs, sequential presentation of four shapes; last dashed line, offset of fixation point. If presented shapes in consecutive epochs are strongly predictive of an alternative, the activity of the population selective for that alternative increases, whereas the competing population is suppressed. (**b**) Sample trials where the same shape, with WOE = 0.7, is presented in the first three epochs; in the last epoch, any of the three shapes with WOE ≥ 0.7 is presented. (**c**) Sample trials where the same shape, with WOE = 0.3, is presented in the first three epochs; in the last epoch, any shape with WOE ≥0.3 is presented. In this case, the population *B* wins the competition and determines the choice of the network on some trials. If presented shapes are not strongly predictive of either alternative, the difference between the activities of two populations remains small during the trial.

no bias (**Supplementary Fig. 8**; but see **Supplementary Fig. 9** for alternative Bayesian observers that show bias in the choice behavior).

To elucidate the bias in the choice behavior, we examined how the synapses learn about evidence related to each shape in this modified task. We found that the difference in the synaptic strengths is increased with the prior, while it is still a linear function of the log naive posterior odds (**Fig. 6c**), with a nearly identical slope (hence independent of the prior). Notably, the difference in the synaptic strengths could be fitted as a linear function of the log naive posterior and log prior odds:

$$\Delta c_i^{ss} = \alpha \log_{10}\left(\frac{\tilde{P}(A\,|\,S_i)}{\tilde{P}(B\,|\,S_i)}\right) + \beta \log_{10}\left(\frac{P(A)}{P(B)}\right) \tag{7}$$

where linear fitting yielded $\alpha = 0.48$ and $\beta = -0.31$.

alternatives, these data can equivalently be expressed in terms of posterior odds. In our model, evidence from multiple cues is combined by effectively summing the log naive posterior odds, which are different from log likelihood ratios when priors are not equal. Therefore, we next explored the model's behavior in simulations when the priors were not equal (see Online Methods for details).

The model makes decisions based on the differential input, which computes the sum of the log naive posterior odds of shapes in a given pattern. The latter is proportional to the log posterior odds for that pattern (**Supplementary Fig. 7a**), a general quantity that we used to express the psychometric function (**Fig. 6a**).

It is evident that the model's choice behavior is strongly biased toward the more probable alternative (that is, the alternative that is assigned a reward more often—*A* in these simulations). We defined the bias in the choice behavior to be the probability of choosing *A* when the log posterior odds is zero. We found the bias in the choice behavior of our model to monotonically increase with the log prior odds (**Fig. 6b**). In contrast, an ideal Bayesian observer would display
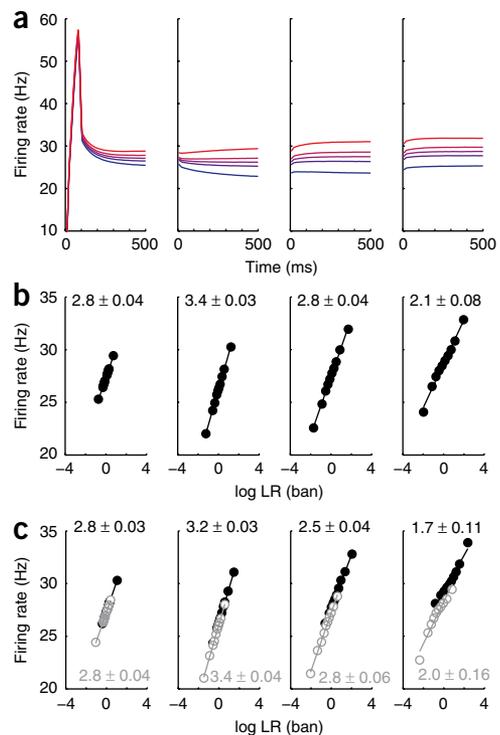
Figure 5 Neural population activity is parametrically correlated with the log LR. (**a**) Effect of the log LR on the firing rate of decision neurons. The population activity in each epoch is aligned on the onset of each shape stimulus presentation, and the average over many trials is computed for five quintiles of the log LR in that epoch (plotted in different colors; more red means larger log LR favoring alternative *A*). The log LR in each epoch is equal to the sum of the log LR of shapes that are presented before and during that epoch. (**b**) Average population firing rate as a function of the log LR for four epochs. Average firing rate during the last 250 ms is computed by grouping the log LR (in base 10, or the unit called a 'ban') into ten equal bins in each epoch. (**c**) Average activity as a function of the log LR in each epoch, plotted separately for trials where the choice is the preferred (black) or nonpreferred (gray) of the neural population. Slope with estimated s.e.m. is shown for each linear fit separately.
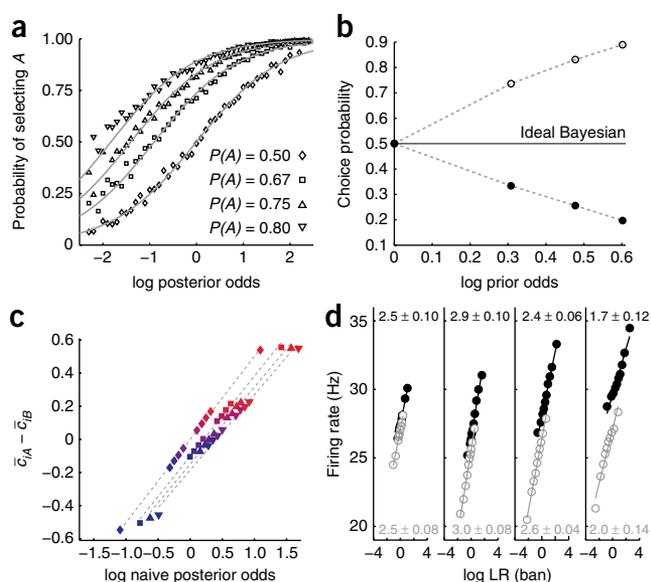
Figure 6 Effect of prior probability on the choice behavior and neural activity. (**a**) Psychometric function for patterns with finite log posterior odds, plotted for four values of prior probability. Gray lines, logistic function fits. (**b**) Bias of the psychometric function (open circles) and of the differential synaptic input (learned evidence) about each shape (filled circles), as a function of the log prior odds. These biases are defined as the probability of selecting choice *A* at zero log posterior odds. An ideal Bayesian observer shows no bias. The dashed lines are only to guide the eyes. (**c**) The difference in the synaptic strengths for each shape as a function of the log naive posterior odds for different values of prior probability. Symbols same as in **a**; dashed lines, linear fits. Different colors represent different shapes. (**d**) Average population activity in each epoch as a function of the log LR in that epoch in the case of *P*(*A*) = 0.8. The black (or gray, respectively) points show data from trials where the choice is the preferred (or nonpreferred, respectively) of the neural population. Slope with estimated s.e.m. is shown for each linear fit separately.

Hence, there is also a bias in the information stored in plastic synapses concerning the evidence provided by each shape stimulus, when priors are not equal. Let us define the bias in the learned evidence as the probability of choosing *A* when one shape is presented alone (after learning) and when the log naive posterior odds is zero (note that there is no shape with zero log naive posterior odds, so this measure of bias is based on extrapolation). This probability (that is, choice bias) is less than 0.5 and is proportional to the log prior odds (**Fig. 6b**; see equation (9) in **Supplementary Note 2**). Therefore, if, after learning, the model is asked to make a judgment based on a single cue, the choice behavior is biased toward the less probable alternative (*B* in this case). At first sight, this result would seem to contradict the observed bias in the psychometric function toward the more probable alternative, but it can be explained as follows (for complete explanation, see **Supplementary Note 2**).

When a pattern consisting of four shapes is presented, the sum of the log naive posterior odds of shapes in that pattern is proportional to the log posterior odds for that pattern but is also positively biased by the log prior odds (**Supplementary Fig. 7a**):

$$\sum_i \log_{10}\left(\frac{\tilde{P}(A\,|\,S_i)}{\tilde{P}(B\,|\,S_i)}\right) = \gamma \log_{10}\left(\frac{P(A\,|\,C^t)}{P(B\,|\,C^t)}\right) + \lambda \log_{10}\left(\frac{P(A)}{P(B)}\right) \quad (8)$$

where $P(A|C^t)$ is the posterior probability that *A* is assigned a reward given that a set of patterns C^t is presented, $\gamma = 0.36$, and $\lambda = 3.64$. The sum of the log naive posterior odds provides an estimate of the log posterior odds because the reward assignment is based on the sum of the WOE of shapes in each pattern. Combined with equation (7), we see that the total differential synaptic inputs $\sum_i \Delta c_i^{ss}$, which determines the choice behavior of the model, is given by

$$\sum_i \Delta c_i^{ss} = \alpha\gamma \log_{10}\left(\frac{P(A\,|\,C^t)}{P(B\,|\,C^t)}\right) + (\alpha\lambda + 4\beta)\log_{10}\left(\frac{P(A)}{P(B)}\right) \quad (9)$$

At zero posterior odds, because $(\alpha\lambda + 4\beta) > 0$, the overall effect of prior is positive and the choice behavior is biased toward the more probable alternative (**Fig. 6a,b**). These results are robust and are not sensitive to the model parameters (see **Supplementary Fig. 10** and **Notes 4** and **5**).

Intuitively, with increasing prior, plastic synapses onto decision neurons selective for the more probable alternative (*A*) are potentiated

more often, and the difference in the synaptic strengths for all shapes becomes more positive. However, according to our learning rule, synapses are updated collectively on each trial, independently of their exact roles in the ultimate decision, and the prior information is mixed with, and attenuated by, the evidence provided by different sets of synapses. Therefore, the influence of the prior on each set is smaller than it should be, and this results in a bias toward the less probable alternative in the estimate of predictive power of each shape. By contrast, when four shapes are presented together, the influence of the prior on each of four shapes in a pattern adds up and amounts to a bias toward the more probable alternative.

We have shown that the differential synaptic input is linearly proportional to the log naive posterior odds, with a slope approximately independent of the prior probability (**Fig. 6c**). Consequently, the average neural activity in each epoch (dictated by the differential synaptic input) depends linearly on the log LR in that epoch, and the slope is only weakly influenced by the prior probability (**Fig. 6d**). The effect of the prior, however, is manifested in the range of firing rates, as the prior induces a shift in the differential synaptic input. As can be seen by comparing **Figure 5c** with *P*(*A*) = 0.5 and **Figure 6d** with *P*(*A*) = 0.8, the firing rates are more markedly different with a larger prior, when the choice is the preferred or nonpreferred target of the decision neurons. This difference in the neural activity gives rise to biased choice behavior toward the alternative with a larger prior.

DISCUSSION
The main findings of this paper are threefold. First, summing log posterior odds, a seemingly complicated calculation, can be readily realized, through approximations, by a plausible plasticity mechanism with bounded synapses in a decision circuit. Second, a biophysically based neural circuit model implementation of the monkey weather-prediction task[6] quantitatively accounted for many behavioral and single-unit neurophysiological observations with a small number (3) of free parameters. Third, considering situations wherein the choice alternatives have unequal priors led us to nontrivial predictions about deviations from the Bayes decision rule.

Inference and combination of information
The weather prediction task exemplifies complex decision making in which one must acquire information concerning the predictive power of each sensory cue as well as combine evidence from multiple cues to make a choice. In this work, we showed that a decision neural circuit model endowed with a simple form of reward-dependent synaptic plasticity is capable of such probabilistic reasoning. Hence, such a high-level cognitive function may

be instantiated by reward-dependent learning, rather than by the sophisticated strategies assumed in some human studies[23].

We showed that plastic synapses in our model dynamically learn and store the association between each shape and outcome in just a few hundred trials, despite the large number of patterns in this task (10^4 patterns). This happens because a synaptic plasticity rule that assumes independence between sources of information enables the system to learn regularities in the external world quickly and robustly. This rule also allows plastic synapses to encode the predictive power related to each shape in the form of the naive posterior probability. As a result, the predictive weight assigned by the model to each shape is smaller than the assigned weight of evidence, as observed experimentally. Note that when conditioned on an outcome (for example, target A is assigned a reward), the evidence samples provided by different cues are no longer statistically independent. Such conditional dependence is complex and may be ignored by the brain in the weather prediction task[6]. We propose that synaptic computation of the naive posteriors, which do not take into account conditional dependences, provides a simple cellular mechanism for the brain to perform inference and cue combination.

Decision neurons integrate evidence from different cues, simply through convergence of synaptic inputs from value-encoding neurons. The strong recurrent dynamics of the decision circuit is critical for generating choices stochastically on single trials; the trial-averaged probability of choosing an option is a sigmoid function of the difference in the inputs associated for each choice option[9–11]. The latter, as we have shown, is approximately proportional to the sum of the log naive posterior odds. Therefore, by summing the log naive posterior odds of shapes in a pattern, our model uses a different strategy than ideal Bayesian observers. Nevertheless, its choice behavior is close to the probabilistic Bayesian observer who follows 'probability matching'[24–26], and so it provides a biophysical instantiation of the probabilistic Bayesian decision rule for two-alternative choice tasks while using a conceptually different framework.

Influence of prior information and 'base-rate neglect'

To differentiate the effect of log likelihood ratios and log posterior odds on decision making, we simulated the weather prediction task with unequal prior probabilities. We found that the model's choice behavior is biased toward the more probable alternative, whereas the information encoded by plastic synapses concerning each shape is biased toward the less probable alternative. These predictions are supported by evidence from the weather prediction task in a human study in which subjects predict one of the two outcomes (for example, rain or sunshine) after observing one, two, three or four tarot cards[1,20]. At the end of the experiment, the subjects are asked to estimate the strength of association between a card and an outcome. When the prior probabilities are not equal, a card that is equally predictive of each outcome is perceived to be more predictive of the less probable outcome, a phenomenon that is known as base-rate neglect and has been described as a judgment fallacy[27,28]. At the same time, the choice behavior is biased in favor of the more probable alternative (see Table 1 in ref. 20).

Using our model, we can explain these counterintuitive results in terms of a plausible biophysical mechanism. Although a few models have been proposed to explain base-rate neglect[20,29], all these mechanistic models assume learning mechanisms that require access to all connection weights in the network and, moreover, do not pertain to any biophysical mechanisms and constraints. Our model prediction on combination of information from different sources through addition of the log naive posterior odds can be tested more directly experimentally. For instance, if after learning the subject must predict

the outcome of a number of cues together (for example, one, two or three), we expect that the bias in these predictions will be proportional to $\beta \log_{10}\left(\dfrac{P(A)}{P(B)}\right)$ times the number of cues (equation (9)).

LIP neural activity and probability representation

Various oculomotor experiments in monkeys have shown that activity of LIP neurons encodes decision variables[8] and is correlated with reward values of choices[19,30–33]. It has been proposed theoretically that a neural population, such as that in LIP, can represent probability distributions concerning sensory information on each trial, which in turn can be used to perform optimal decision making[34–36]. Adding to this body of literature, a recent experiment demonstrated that activity of LIP neurons reflects probability integration, namely the summation of the log LR (ref. 6). This quantity, however, can be only computed by tabulating the frequency of occurrence of each shape combination and the outcome in different epochs of the task (see Online Methods). Our model suggests that LIP neurons may reflect reward probability, such as log posterior odds (rather than log LR); but posteriors are encoded at synapses onto action value–encoding neurons that project to LIP. Consequently, on every trial, LIP neurons integrate reward information and contribute to decision making. Our model prediction can be tested experimentally using unequal priors, which would make it possible to differentiate the log naive posterior odds and the log LR.

The plastic synapses proposed in our model should be found in neural circuits involved with representation of stimulus–reward or action–reward associations, such as parts of the prefrontal cortex[37,38] and basal ganglia[39–41]. Moreover, corticostriatal synapses show long-term potentiation and depression that depend on the presence or absence of dopamine modulation[42,43], and dopamine neurons are involved in reward signaling[44,45]. All together, these findings provide strong neurophysiological support for the synaptic plasticity rule used in this paper. Consistently, using functional brain imaging in humans, it has been shown that the striatum gradually becomes active as learning progresses in the weather prediction task[46]. It would be worthwhile, in future experiments, to test whether this and other brain areas encode reward probabilities in the form of log posterior odds.

Our model is general and can be applied to different probabilistic decision making tasks. Indeed, we have used a similar learning rule and decision making mechanism to capture a foraging behavior known as the matching law[9], as well as choice behavior in a competitive game[10]. This work shows how complicated inference and cue combination can be performed by a recurrent decision circuit endowed with a plausible synaptic plasticity rule. Perhaps other high-level cognitive abilities can be instantiated by simple neural mechanisms as well.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

AUTHOR CONTRIBUTIONS
A.S. and X.-J.W. conceived the problem and designed the model. A.S. performed model simulations and analyzed the data. A.S. and X.-J.W. wrote the paper.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Knowlton, B.J., Squire, L.R. & Gluck, M.A. Probabilistic classification learning in amnesia. *Learn. Mem.* **1**, 106–120 (1994).
2. Knowlton, B.J., Mangels, J.A. & Squire, L.R. A neostriatal habit learning system in humans. *Science* **273**, 1399–1402 (1996).
3. Moody, T.D., Bookheimer, S.Y., Vanek, Z. & Knowlton, B.J. An implicit learning task activates medial temporal lobe in patients with Parkinson's disease. *Behav. Neurosci.* **118**, 438–442 (2004).
4. Fera, F. *et al.* Neural mechanisms underlying probabilistic category learning in normal aging. *J. Neurosci.* **25**, 11340–11348 (2005).
5. Ashby, F.G. & Maddox, W.T. Human category learning. *Annu. Rev. Psychol.* **56**, 149–178 (2005).
6. Yang, T. & Shadlen, M.N. Probabilistic reasoning by neurons. *Nature* **447**, 1075–1080 (2007).
7. Gold, J.I. & Shadlen, M. Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* **5**, 10–16 (2001).
8. Gold, J.I. & Shadlen, M.N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
9. Soltani, A. & Wang, X.-J. A biophysically-based neural model of matching law behavior: melioration by stochastic synapses. *J. Neurosci.* **26**, 3731–3744 (2006).
10. Soltani, A., Lee, D. & Wang, X.-J. Neural mechanism for stochastic behavior during a competitive game. *Neural Netw.* **19**, 1075–1090 (2006).
11. Fusi, S., Asaad, W.F., Miller, E.K. & Wang, X.-J. A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple timescales. *Neuron* **54**, 319–333 (2007).
12. Fusi, S., Drew, P.J. & Abbott, L.F. Cascade models of synaptically stored memories. *Neuron* **45**, 599–611 (2005).
13. Fusi, S. & Abbott, L.F. Limits on the memory storage capacity of bounded synapses. *Nat. Neurosci.* **10**, 485–493 (2007).
14. Wang, X.-J. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* **36**, 955–968 (2002).
15. Wong, K.-F. & Wang, X.-J. A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.* **26**, 1314–1328 (2006).
16. Wong, K.-F., Huk, A.C., Shadlen, M.N. & Wang, X.-J. Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making. *Front. Comput. Neurosci.* **1**, 6 (2007).
17. Furman, M. & Wang, X.-J. Similarity effect and optimal control of multiple-choice decision making. *Neuron* **60**, 1153–1168 (2008).
18. Liu, F. & Wang, X.-J. A common cortical circuit mechanism for perceptual categorical discrimination and veridical judgment. *PLOS Comput. Biol.* **4**, e1000253 (2008).
19. Wang, X.-J. Decision making in recurrent neuronal circuits. *Neuron* **60**, 215–234 (2008).
20. Gluck, M.A. & Bower, G.H. From conditioning to category learning: an adaptive network model. *J. Exp. Psychol. Gen.* **117**, 227–247 (1988).
21. Amit, D.J. & Fusi, S. Dynamic learning in neural networks with material synapses. *Neural Comput.* **6**, 957–982 (1994).
22. Fusi, S. Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. *Biol. Cybern.* **87**, 459–470 (2002).
23. Meeter, M., Myers, C.E., Shohamy, D., Hopkins, R.O. & Gluck, M.A. Strategies in probabilistic categorization: results from a new way of analyzing performance. *Learn. Mem.* **13**, 230–239 (2006).
24. Myers, J.L. Probability learning and sequence learning. in *Handbook of Learning and Cognitive Processes* (ed. Estes, W.K.) 171–205 (Erlbaum, Hillsdale, New Jersey, USA, 1976).
25. Vulkan, N. An economist's perspective on probability matching. *J. Econ. Surv.* **14**, 101–118 (2000).
26. Shanks, D.R., Tunney, R.J. & McCarthy, J.D. A re-examination of probability matching and rational choice. *J. Behav. Decis. Mak.* **15**, 233–250 (2002).
27. Kahneman, D. & Tversky, A. On the psychology of prediction. *Psychol. Rev.* **80**, 237–251 (1973).
28. Tversky, A. & Kahneman, D. Evidential impact of base rates. in *Judgment Under Uncertainty: Heuristics and Biases* (eds. Kahneman, D., Slovic, P. & Tversky, A.) 153–160 (Cambridge Univ. Press, Cambridge, UK, 1982).
29. Kruschke, J.K. ALCOVE: an exemplar-based connectionist model of category learning. *Psychol. Rev.* **99**, 22–44 (1992).
30. Glimcher, P.W. The neurobiology of visual-saccadic decision making. *Annu. Rev. Neurosci.* **26**, 133–179 (2003).
31. Sugrue, L.P., Corrado, G.C. & Newsome, W.T. Matching behavior and representation of value in parietal cortex. *Science* **304**, 1782–1787 (2004).
32. Sugrue, L.P., Corrado, G.S. & Newsome, W.T. Choosing the greater of two goods: neural currencies for valuation and decision making. *Nat. Rev. Neurosci.* **6**, 363–375 (2005).
33. Soltani, A. & Wang, X.-J. From biophysics to cognition: reward-dependent adaptive choice behavior. *Curr. Opin. Neurobiol.* **18**, 209–216 (2008).
34. Ma, W.J., Beck, J.M., Latham, P.E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* **9**, 1432–1438 (2006).
35. Beck, J.M. *et al.* Probabilistic population codes for bayesian decision making. *Neuron* **60**, 1142–1152 (2008).
36. Ma, W.J., Beck, J.M. & Pouget, A. Spiking networks for bayesian inference and choice. *Curr. Opin. Neurobiol.* **18**, 217–222 (2008).
37. Rushworth, M.F.S. & Behrens, T.E.J. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* **11**, 389–397 (2008).
38. Lee, K.-M. & Keller, E.L. Neural activity in the frontal eye fields modulated by the number of alternatives in target choice. *J. Neurosci.* **28**, 2242–2251 (2008).
39. Lauwereyns, J., Watanabe, K., Coe, B. & Hikosaka, O. A neural correlate of response bias in monkey caudate nucleus. *Nature* **418**, 413–417 (2002).
40. Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of action-specific reward values in the striatum. *Science* **310**, 1337–1340 (2005).
41. Lau, B. & Glimcher, P.W. Value representations in the primate striatum during matching behavior. *Neuron* **58**, 451–463 (2008).
42. Reynolds, J.N., Hyland, B.I. & Wickens, J.R. A cellular mechanism of reward-related learning. *Nature* **413**, 67–70 (2001).
43. Shen, W., Flajolet, M., Greengard, P. & Surmeier, D.J. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* **321**, 848–851 (2008).
44. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
45. Schultz, W. Multiple dopamine functions at different time courses. *Annu. Rev. Neurosci.* **30**, 259–288 (2007).
46. Poldrack, R.A. *et al.* Interactive memory systems in the human brain. *Nature* **414**, 546–550 (2001).

ONLINE METHODS

Description of the weather prediction task. In the simulated experiment, monkeys were trained to choose between two color targets (green and red) after observing four shapes that were presented on a screen sequentially at 500-ms intervals[6]. These shapes were selected randomly from a set of ten distinguishable shapes with replacement, $\{S_1, S_2, ..., S_{10}\}$, each of which was assigned a unique weight of evidence (WOE). The WOE for each shape was defined as the log LR that a red or green target was assigned a reward; or, equivalently, the selection of red or green target was accompanied by a reward. The WOEs for these ten shapes, $\{w_1, w_2, ..., w_{10}\}$, were chosen to be $[-\infty, -0.9, -0.7, -0.5, -0.3, 0.3, 0.5, 0.7, 0.9, \infty]$ in favor of the red target. For example, the presentation of a shape with WOE = 0.9 by itself predicted that the red target was assigned a reward 89% $\left(= \dfrac{10^{0.9}}{1 + 10^{0.9}}\right)$ of the times.

On each trial, four shapes (which we call a pattern) were presented on the screen and either of the two targets was assigned a reward with a probability depending on the sum of the WOEs of all shapes in that pattern. More specifically, the probability that the red target was assigned a reward given that four shapes were presented was equal to

$$P(R \mid s_1, s_2, s_3, s_4) = \frac{10^{\sum_{n=1}^{4} w_n}}{1 + 10^{\sum_{n=1}^{4} w_n}} \tag{10}$$

where s_n represents the shape shown in the nth epoch. The probability that the green target was assigned a reward was $1 - P(R \mid s_1, s_2, s_3, s_4)$.

To introduce an unequal prior, we first generated a set of patterns according to the paradigm described and then randomly removed a portion of trials in which the reward was assigned to the less probable alternative. This alteration changes the prior probability that each alternative is rewarded without changing the structure of the task.

Description of the model and learning rule. The model is an extended version of our previous biophysically-based model of probabilistic decision-making network[9,10,14]. The decision making circuit of the model is a firing-rate model which has been shown to reproduce the choice and neural activity of the detailed spiking network model[15]. All details about the decision circuit of the model and its parameters were reported elsewhere[15].

The model consists of three layers (**Fig. 1a**). The first layer contains sensory neurons that are selective for visual cues (shapes). These cue-selective neurons can be located in the inferotemporal cortex, which has been shown to contain neurons that encode different shapes by a combination of active and inactive columns selective for individual features[47,48]. The cue-selective neurons project to the second layer, where neurons learn to encode reward values of the two alternative responses (action values) through plastic synapses that undergo reward-dependent Hebbian modifications[9–11] (see below). The second layer therefore presumably corresponds to certain frontal areas, such as the anterior cingulate cortex or the dorsolateral prefrontal cortex, that are known to be involved with representation and learning of action values[37,38]. The convergence from cue-selective neurons enables value-encoding neurons to combine information from different cues. The third layer is a decision circuit with two competing neural pools that are selective for choice ('preferred target') A and B, respectively. This decision circuit was modeled in the same way as in our previous work[9,14,15]. We compared the firing activity of these decision neurons with neural data recorded from LIP[6].

Computer simulations followed the experimental protocol. On a simulated trial, at the onset of the first shape stimulus, the visual inputs (representing fixation point as well as the visual cue) triggered a brief transient response of neurons in the decision-making network that decreased to a moderate level, similarly to LIP neurons[6]. Upon the presentation of each shape, the activity of sensory neurons selective for that shape was increased from zero to a constant value and this activity was sustained throughout the trial (if a shape was repeated on a trial, the activity of the corresponding population was multiplied by the number of repetitions of the shape on that trial). At the end of the trial, when the fixation point goes off, the activity of these two populations drops due to a decrease in the overall inputs (see below). This is followed by a divergence of activity between the two populations that, as a distinctive feature of competition

in the decision-making network, signals the choice of the model on each trial[15]. Specifically, the model's decision is determined by the neural population that is the first to reach a fixed firing rate threshold of 30 Hz. In general, a population of neurons receiving larger inputs reaches a higher level of activity and consequently has a higher chance to win the competition and determines the model's choice on a trial.

Because the sensory responses of cue-selective neurons are similar, the only factor that differentiates the inputs to decision neurons is the strength of plastic synapses from sensory neurons onto value-encoding neurons. In addition to these inputs, decision neurons also receive a large background input as well as purely visual inputs, which mimics the visual response of neurons in the visual cortex and keeps the decision circuit from entering the competition regime during the presentation of four shapes and before the extinction of fixation point (see **Supplementary Fig. 11** and **Supplementary Methods** for more details).

The inputs to decision neurons are determined by the firing activity of sensory neural populations encoding the presented shapes, and by the strength of plastic synapses from these populations onto value-encoding populations. We assumed these plastic synapses are binary (that is, they only have two stable states)[21,22], on the basis of evidence that plastic synapses have discrete (binary) states[49,50]. Here we used binary synapses, but our results still hold with multiple discrete states. For binary synapses, the average strength of these synapses can be defined as the fraction of synapses in the potentiated state, denoted by c_{iA} and c_{iB} (for synapses from neurons selective for shape S_i onto value-encoding neurons selective for alternative A or B, respectively).

At the end of each trial, plastic synapse were modified according to a stochastic reward-dependent Hebbian learning rule[9–11,21,22]. First, Hebbian plasticity requires high activity in both pre- and postsynaptic neurons, so only synapses from cue-selective neurons selective for presented shapes onto the value-encoding neurons selective for the chosen alternative were modified. Second, depending on the outcome (reward or no reward) of a given trial, plastic synapses were potentiated or depressed. Third, these modifications took place stochastically.

At the end of each trial, only neural populations selective for all presented shapes (through working memory) and for the selected alternative were active. If the choice of the model was rewarded, all sets of synapses selective for presented shapes onto the value-encoding neurons selective for the chosen alternative (say A) were potentiated, with probability q_+. That is, all synapses in the depressed state made a transition to the potentiated state with probability q_+. As a result, the synaptic strength for each set of plastic synapses (say i) was updated as follows

$$c_{iA} \rightarrow c_{iA} + q_+(1 - c_{iA}) \tag{11}$$

Alternatively if the choice of the model was not rewarded, plastic synapses were depressed with probability q_-, so the synaptic strength was updated as

$$c_{iA} \rightarrow c_{iA} - q_- c_{iA} \tag{12}$$

For all simulations presented in the paper, we set $q_+ = 0.02$ and $q_- = 0.02$, except for the analytical results in **Fig. 1** and **Supplementary Note 4** where we show how the model's choice of behavior depends on the learning rates.

Data analysis. All of the data analysis and average values reported here were computed over 100,000 trials of the simulated experiment, except values related to the neural activity of decision circuit, which were computed using 20,000 simulated trials. To estimate the influence of each shape on the model's choice behavior, we used a logistic regression fit similar to one used for the monkey experiment[6]. We assumed that probability of selecting a choice is influenced by the presence of each shape as

$$P_A = \frac{10^Q}{1 + 10^Q} \text{ where } Q = \sum_{i=1}^{10} q_i N_i \tag{13}$$

where the N_i is the number of appearances of shape i in each pattern. The regression coefficients, q_i, are called the subjective weight of evidence (SWOE) and measure the influence of each shape on decision making.

We used the log LR through the paper as a measure of evidence provided by the presented shapes until a certain epoch or the end of a trial. The log LR that a target is assigned a reward is equal to

$$\log \mathrm{LR}_n = \log_{10} \frac{P(s_1,...,s_n \mid \text{reward at } A)}{P(s_1,...,s_n \mid \text{reward at } B)} \quad n = 1,2,3,4 \qquad (14)$$

This quantity was computed by tabulating the frequency of reward being assigned to each shape combination for each epoch. To compute the change in the log LR due to presentation of each shape, we calculated the average change in the log LR due to the presentation of another shape from one epoch to the next, while excluding the trump shapes if they predicted one of the outcomes deterministically (similarly to ref. 6).

47. Tanaka, K. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
48. Tsunoda, K., Yamane, Y., Nishizaki, M. & Tanifuji, M. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat. Neurosci.* **4**, 832–838 (2001).
49. Petersen, C.C., Malenka, R.C., Nicoll, R.A. & Hopfield, J.J. All-or-none potentiation at CA3–CA1 synapses. *Proc. Natl. Acad. Sci. USA* **95**, 4732–4737 (1998).
50. O'Connor, D.H., Wittenberg, G.M. & Wang, S.S.-H. Graded bidirectional synaptic plasticity is composed of switch-like unitary events. *Proc. Natl. Acad. Sci. USA* **102**, 9679–9684 (2005).