nature neuroscience

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).

Received 12 May; accepted 5 October; published online 13 December 2009; doi:10.1038/nn.2450

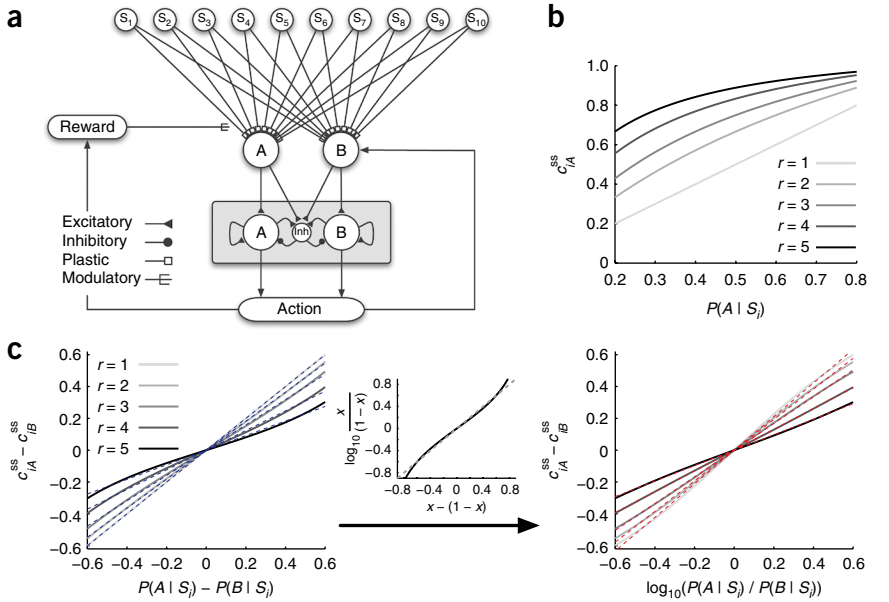

Figure 1 Schematic of the model and posterior computation by plastic synapses when a single cue is presented on each trial. (**a**) Schematic of the three-layer model. The first layer consists of cue-selective neural populations, each is activated upon the presentation of a cue (a shape in the weather 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

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

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

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

This formula can be simplified by observing that the second term in the denominator, $k = (r-1)^2 P(A|S_i)P(B|S_i)$, 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 $0.2 \le P(A|S_i) \le 0.8$, $4(r-1)^2 / 25 \le k \le (r-1)^2 / 4$). Since *k* is roughly constant, the difference in the steady state of the synaptic

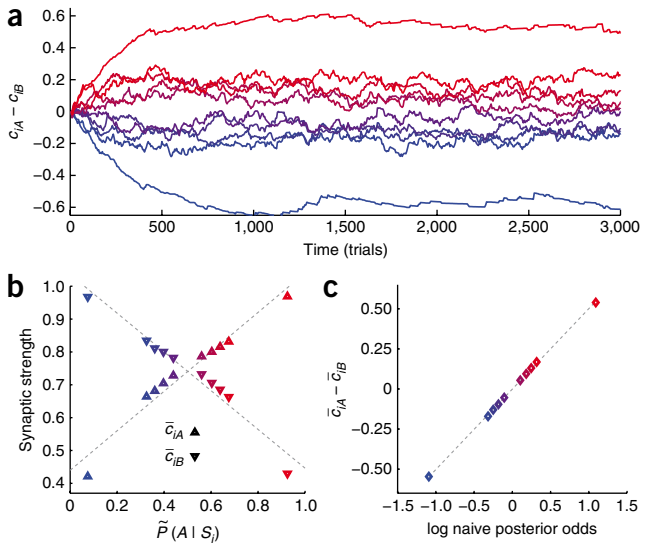

Figure 2 Posterior computation by plastic synapses when multiple cues are presented on each trial. (**a**) Time course of learning in plastic synapses. Shown is the difference in the strengths of synapses from a cue-selective neural population onto the two action value–encoding neural populations (*A* 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**)

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

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**):

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

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 $P(A \mid S_i)$ (**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

$$\mathrm{WOE} = \log_{10} \frac{P(A \mid S_i)}{P(B \mid S_i)}$$

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, $\tilde{P}(A \mid S_i)$, 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**)

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

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

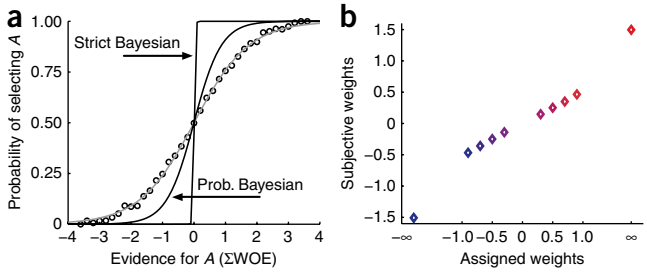

Figure 3 Choice behavior of the model and the subjective weight of evidence in the weather prediction task. (**a**) Probability of choosing alternative *A* as a function of the evidence favoring this alternative for all patterns with finite WOEs. The evidence is equal to the sum of the WOE of all shapes 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):

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

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

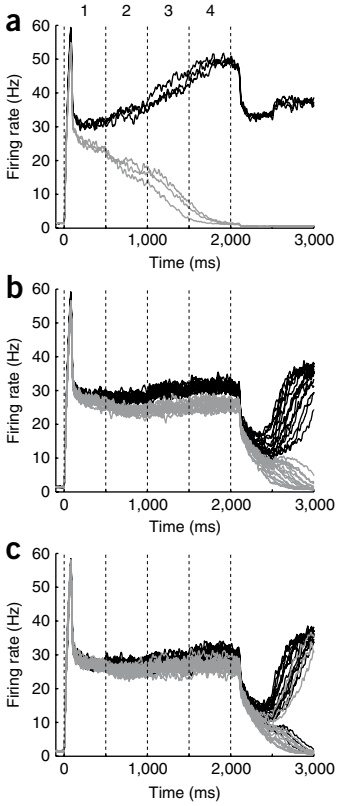

Figure 4 Model neural population activity during the weather prediction task. (**a**) Firing activity of two (black, *A*; gray, *B*) choice-selective populations in the decision-making network on a few sample trials. On these trials, the trump shape favoring *A* (with infinite WOE) is presented in epochs 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.

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 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:

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

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

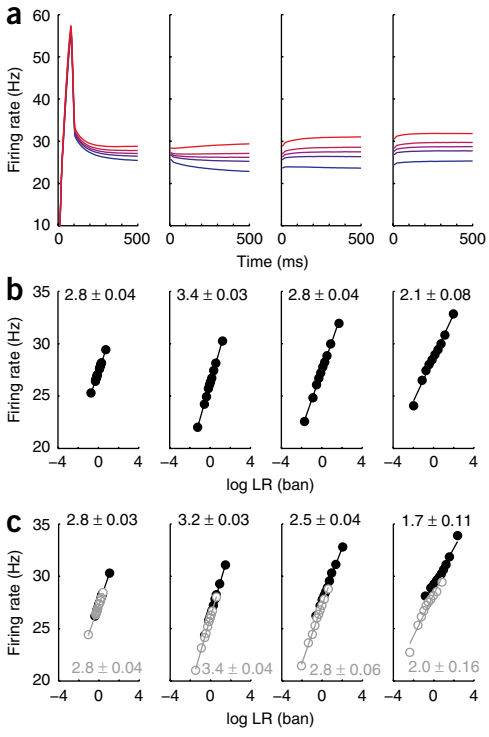

Figure 5 Neural population activity is parametrically correlated with the log LR. (**a**) Effect of the log LR on the firing rate of decision neurons. The population activity in each epoch is aligned on the onset of each shape stimulus presentation, and the average over many trials is computed for five 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.

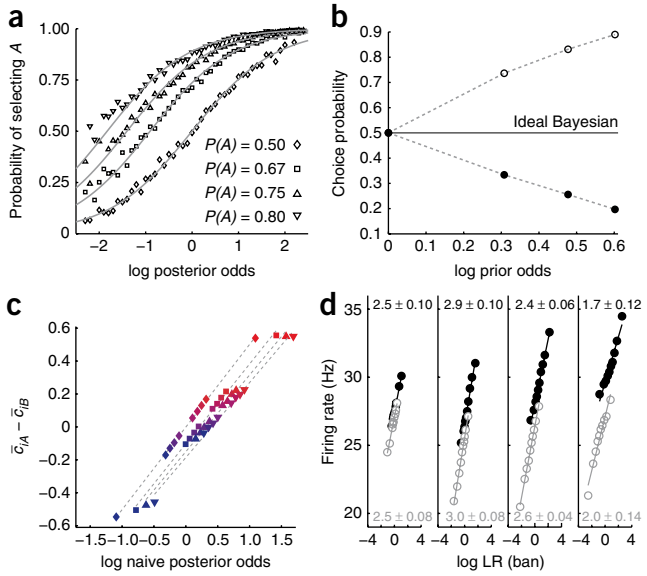

Figure 6 Effect of prior probability on the choice behavior and neural activity. (**a**) Psychometric function for patterns with finite log posterior odds, plotted for four values of prior probability. Gray lines, logistic function fits. (**b**) Bias of the psychometric function (open circles) and of 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 $\beta \log_{10}\left(\frac{P(A)}{P(B)}\right)$ 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.

ACKNOWLEDGMENTS

This work was supported by US National Institutes of Health grants 2-R01-MH062349 and MH073246. We are thankful to D. Andrieux, S. Ardid, A. Bernacchia and R. Wilson for comments on the manuscript.

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.Aneural 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, $\{S_1, S_2, \ldots, S_{10}\}$, 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, $\{w_1, w_2, \ldots, w_{10}\}$, 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% $\left(=\frac{10^{0.9}}{1+10^{0.9}}\right)$ 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

doi:10.1038/nn.2450

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

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).

Behavioral/Systems/Cognitive

Visual Saliency Computations: Mechanisms, Constraints, and the Effect of Feedback

Alireza Soltani[1,2] **and Christof Koch**[1]

[1]Division of Biology and Computation and Neural Systems, California Institute of Technology, Pasadena, California 91125, and [2]Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030

The primate visual system continuously selects spatial proscribed regions, features or objects for further processing. These selection mechanisms— collectively termed selective visual attention—are guided by intrinsic, bottom-up and by task-dependent, top-down signals. While much psychophysical research has shown that overt and covert attention is partially allocated based on saliency-driven exogenous signals, it is unclear how this is accomplished at the neuronal level. Recent electrophysiological experiments in monkeys point to the gradual emergence of saliency signals when ascending the dorsal visual stream and to the influence of top-down attention on these signals. To elucidate the neural mechanisms underlying these observations, we construct a biologically plausible network of spiking neurons to simulate the formation of saliency signals in different cortical areas. We find that saliency signals are rapidly generated through lateral excitation and inhibition in successive layers of neural populations selective to a single feature. These signals can be improved by feedback from a higher cortical area that represents a saliency map. In addition, we show how top-down attention can affect the saliency signals by disrupting this feedback through its action on the saliency map. While we find that saliency computations require dominant slow NMDA currents, the signal rapidly emerges from successive regions of the network. In conclusion, using a detailed spiking network model we find biophysical mechanisms and limitations of saliency computations which can be tested experimentally.

Introduction

A crucial computational strategy of the primate visual system is to swiftly allocate processing resources to a region, feature or object to deal with the many overlapping and partially occluding objects in natural scenes. Attentional selection can be guided by exogenous signals from the environment, such as a red flashing light (bottom-up, saliency-driven attention), by endogenous signals such as when looking for a specific car in a parking lot (top-down, volitional-controlled attention), or by both. Nevertheless, the neural mechanisms underlying these processes are mostly unknown.

From a computational point of view, a purely feedforward model of bottom-up attention incorporating a saliency map successfully predicts a large fraction of fixated locations under free viewing conditions (Itti et al., 1998; Itti and Koch, 2001; Parkhurst et al., 2002; Cerf et al., 2008; Foulsham and Underwood, 2008; Mannan et al., 2009). At its heart there is a two-dimensional (2-D) topographic arrangement of neurons that represent stimulus saliency throughout the visual scene. Initially, prominent locations corresponding to regions with enhanced feature contrast are computed in individual maps (i.e., conspicuity maps) for different dimensions of the stimulus such as intensity, orientation, color, motion, etc. These computations are performed through a set of multiscale, center-surround and normalization operations. Finally, the conspicuity maps are combined to form a single saliency map. Activity in this map does not encode conspicuity in any one particular feature dimension, but encodes the overall conspicuity of a given location relative to its local and global neighborhood. Based on electrophysiological evidence from the monkey, the lateral intraparietal cortex (LIP) and the frontal eye fields (FEF) have been identified as possible saliency maps (Gottlieb et al., 1998; Kusunoki et al., 2000; Bisley and Goldberg, 2003; Moore and Armstrong, 2003; Thompson and Bichot, 2005).

While it is believed that LIP and FEF represent the saliency map, neurons in lower visual areas V1, V4, and V5 (MT) also show differential responses to a target stimulus depending on surrounding stimuli or its spatiotemporal context (Allman et al., 1985; Knierim and van Essen, 1992; Albright and Stoner, 2002; Hegdé and Felleman, 2003; Burrows and Moore, 2009). For example, Hegdé and Felleman (2003) measured the response of V1 neurons to oriented and colored bars in the receptive field (RF) that had different saliency values. In particular, they compared the response to popout targets—say a red bar among green bars that rapidly attracts the eye—to conjunction targets—say a red, vertical bar among red, horizontal and green, vertical and horizontal ones—targets that are defined by the combination of two or more feature dimensions. Such conjunction targets are not readily detectable. They found that V1 neurons do not distinguish between the popout and conjunction targets and therefore, that V1 neurons do not carry saliency signals. More recently, Burrows and Moore (2009) examined the response of V4 neurons to similar stimuli and concluded that these neurons can distinguish between the

Received March 24, 2010; revised July 9, 2010; accepted July 17, 2010.

We are grateful to the Mathers Foundations, the Office of Naval Research, and the Defense Advanced Research Projects Agency for financial support of the research reported here. We thank Zahra Ayubi, Brittany Burrows, and Tirin Moore for helpful discussions and comments on the manuscript.

Correspondence should be addressed to Alireza Soltani at the above address. E-mail: soltani@bcm.edu.

DOI:10.1523/JNEUROSCI.1517-10.2010

 0270-6474/10/3012831-13$15.00/0

popout and conjunction targets. Paradoxically, they also showed that the saliency signal in V4 diminishes if the monkey prepares a saccade to a location far from the RF of the neuron, indicating an important role for top-down attention in the formation of the bottom-up driven saliency signal in V4.

These findings raise some important questions. First, how are saliency signals formed in the visual cortex across the cascade of regions from V1, to V2, V3, V4 and so on? Second, how does top-down attention affect these computations?

To answer above questions and shed light on neural substrates of bottom-up attention and its interaction with top-down attention, we construct a 2-D, biophysically plausible spiking network model. The network contains three distinct layers of neural populations corresponding to three cortical regions—which we identify from here on as V1, V2, and V4—and a higher visual area assumed to instantiate the saliency map (either LIP or FEF). The model neurons receive realistic inputs which are generated from the actual stimuli used in the relevant experiment (Hegdé and Felleman, 2003; Burrows and Moore, 2009). Using our model, we consider biophysical mechanisms and constraints on saliency computations and how these are influenced by top-down attention. Our hypothesis is that feedback from a cortical area representing the saliency map to earlier visual areas improves saliency computations, while top-down attention interferes with these computations by disrupting the feedback through its influence on the activity in the saliency map.

Materials and Methods

We use leaky-integrate-and-fire (LIF) model neurons with realistic synapses as building blocks. Our spiking network model contains many 2-D populations of neurons (24 and 10 in the first and second set of simulations, respectively) with realistic inputs and synapses, making it computationally expensive. For example, simulating 200 trials of the response to a given stimulus takes ~10 h to run on a standard Unix system with a 3 GHz Intel CPU. We therefore had to adopt some simplifications.

First, we assume that inputs to the network are the outputs of lateral geniculate nucleus (LGN) and V1 neurons that are wavelength- and orientation-selective, respectively. We do not explicitly model these cells, using instead their RF properties to generate their response to visual stimuli. As a result, the inputs to the network have wavelength (here the colors red and green) and/or orientation selectivity (0, 45, 90, 135), and we only explicitly model the visual processing in the output layers of V1, cortical areas V2 and V4, and a higher area corresponding to the saliency map (LIP/FEF). Second, we use Cartesian coordinates and ignore the effect of cortical magnification.

For each trial, we simulate 300 ms of the network dynamics with $dt = 0.1$ ms using the improved RK2 integration algorithm (Hansel et al., 1998). We directly compare our model against two different electrophysiological experiments in the alert and behaving monkey (Hegdé and Felleman, 2003; Burrows and Moore, 2009) using the same visual stimuli (see below).

Spiking network model. The model consists of 3 regions of neural populations, each of which contains 4 excitatory and 4 corresponding inhibitory populations (not represented) of LIF units (Fig. 1). Each population consists of 28 × 28 neurons, covering 14° × 14° of the visual field. Therefore, each neuron spans 0.5° of the visual field. We assume periodic boundary conditions for connections between all neurons (i.e., each 2-D neural population is placed on a torus).

We examine two exclusive architectures for the network. In configuration *A* (Fig. 1*A*), individual features (i.e., color and orientation) are processed in separate neural populations, and neurons in different populations receive inputs selective to only one feature. That is, there are no cells which participate in saliency computations and are tuned to both color and orientation. To compare the activity of these model neurons with experimentally recorded neurons selective to two features (say a red bar at 0° orientation), we combine the outputs of neural populations selective to the color red and to 0° orientation (Fig. 1*A*). We formulate this combination by simply adding the spike trains of neurons with similar RF in the corresponding neural populations (to avoid further computations). In configuration *B* (Fig. 1*B*), a combination of features is jointly processed and neurons in different populations receive inputs selective to both orientation and color (e.g., red color and 0° orientation). That is, oriented neurons are also color-tuned and we can directly compare the activity of neurons in this configuration with the experimental data. Apart from the input organization, all parameters are similar for the two configurations.

Model parameters for all excitatory neurons are set to: threshold voltage $V_{th} = -50$ mV, reset voltage $V_{reset} = -55$ mV, leak voltage $V_{leak} = -70$ mV, refractory time period $t_{ref} = 2$ ms, capacitance $C_E = 0.5$ nF, and leak conductance $G_{leak,E} = 25$ nS. Inhibitory interneurons have similar parameters except that the capacitance and leak conductance are set to $C_I = 0.2$ nF and $G_{leak,I} = 20$ nS, respectively. Neurons in each population are connected to all other neurons with a circular Gaussian profile (i.e., with equal width, σ, in both dimensions). That is, most connections are local with synaptic weight falling off with distance.

Excitatory neurons project to their target neurons through two types of synaptic receptors; fast AMPA, with the time constant $\tau_s = 2$ ms, and slow saturating NMDA, with the time constant $\tau_s = 80$ ms. The spatial extent of the connectivity profile (i.e., σ in the Gaussian function) is the same for both receptor types (see supplemental Table 1, available at www.jneurosci.org as supplemental material, for connectivity parameters). Inhibitory neurons are connected to their target neurons through GABA synapses with the time constant $\tau_s = 10$ ms. The peak conductance for all synapses, g_{syn}, is set to 1 nS multiplied by the connection strength (see below). All synapses are modeled as having exponentially decreasing conductances with time.

To capture the observed response adaptation in visual areas, we include spike-rate-adaptation (SRA) current for all neurons. For neurons in feature-selective populations, we set the change in conductance and the time constant of the SRA current to $g_{SRA} = 0.6$ nS and $\tau_{SRA} = 50$ ms, respectively. To reduce the strong response to the single bar stimulus in the saliency population, we adopt a stronger SRA current for neurons in this population ($g_{SRA} = 4.0$ nS and $\tau_{SRA} = 50$ ms).

Every excitatory neuron in a given population is connected to all other neurons in that population and to all interneurons in the corresponding inhibitory population. Cross-orientation inhibition is implemented by subtracting 25% of the mean of all orientation inputs (0°, 45°, 90°, 135°) from a given orientation input. In addition, there are feedforward connections between excitatory neurons with similar feature selectivity in successive regions.

All connection matrices are normalized Gaussian functions, with width σ, multiplied by the weight of these connections, w (see supplemental Table 1, available at www.jneurosci.org as supplemental material). We assume identical values for σ and w in the three simulated regions. Because the connectivity matrices are normalized, the value of w for a given connection should not be taken by itself as the magnitude of that connection strength.

To study the importance of NMDA currents in saliency computations, we reduce NMDA currents while increasing AMPA currents such that their sum remains approximately the same. Specifically, for the two additional sets of simulations presented here, we set the connection strength parameters between excitatory neurons, $w^{AMPA}_{EE,i\to i}$ and $w^{NMDA}_{EE,i\to i}$, equal to [33, 22] and [44, 11], respectively (compare with the original values of [11, 44]) (see supplemental Table 1, available at www.jneurosci.org as supplemental material). As we reduce the NMDA currents between excitatory neurons we also need to reduce the strength of the NMDA currents from the excitatory to inhibitory neurons to avoid the slowly activated inhibitory neurons from suppressing the activity in the network after the onset response. For connections from the excitatory to inhibitory neurons we set $w^{AMPA}_{EI,i\to i}$ and $w^{NMDA}_{EI,i\to i}$, equal to [180, 45] and [202.5, 22.5], respectively (compare with the original values of [135, 90]). Note that the connection widths, σ, are similar for AMPA and NMDA synapses. These and the rest of the model parameters are kept the same for simulations on the role of NMDA receptors.

Finally, to study the effect of feedback from the saliency map on the saliency computations in V4, we use a simplified version of our model

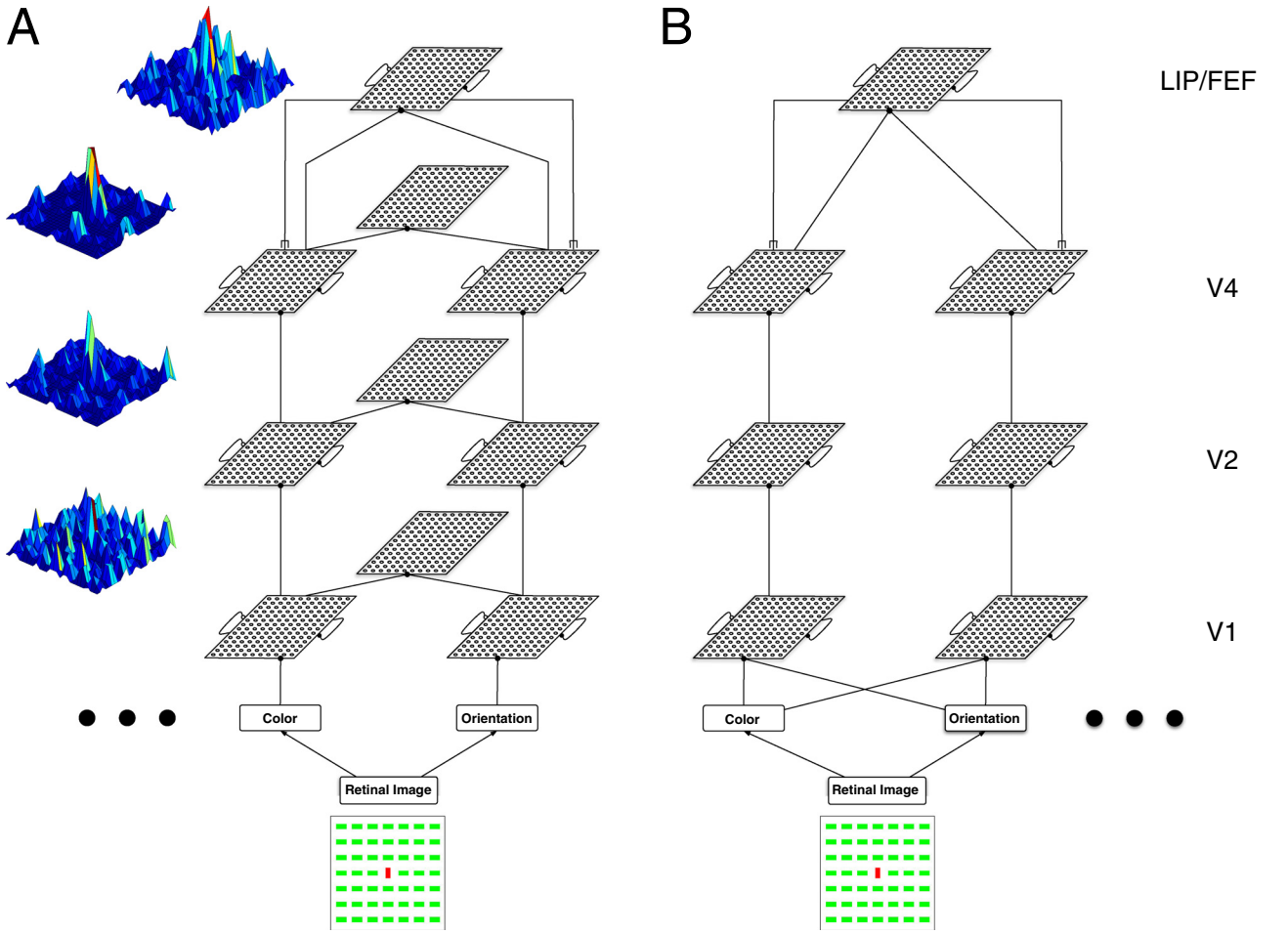

Figure 1. Two alternative architectures of the spiking network model of saliency. The model consists of 3 cortical regions—V1, V2 and V4. Each one includes 8 populations of excitatory and inhibitory interneurons, each modeled by 28 × 28 LIF units. Neurons in each population are connected to all neurons in the same population and to their corresponding interneurons (not represented). The first layer of populations represents neurons in the output layers of V1. Excitatory neurons in V1 (respectively V2) provide excitatory inputs to excitatory neurons at corresponding topographic locations in V2 (respectively, V4). For some simulations, we also included an explicit saliency map that receives inputs from all color and orientation-selective neurons in V4, and provides feedback to both excitatory and inhibitory neurons in V4. ***A***, In configuration *A*, each excitatory population receives an input selective to one of two orientations (0° and 90°) or one of two colors (red and green), and projects to neurons with similar selectivity in the next processing stage. These inputs are generated by the RF properties of LGN and V1 neurons. To compare the results with experimental data, we add the activity of these populations to construct different color- and orientation-selective populations. The insets show the average response (during a sample trial) of the constructed neural populations selective to red-horizontal bar and of the saliency population, to the combined popout display. ***B***, In configuration *B*, each excitatory population receives a combination of inputs selective to orientation and color (red-horizontal, red-vertical, green-horizontal, and green-vertical).

(for the sake of computational efficiency). This simplified model contains two layers of neural populations: the first layer corresponding to feature-selective neurons in V4 and the second layer corresponding to spatially selective saliency neurons in the LIP/FEF. The saliency map consists of one population of excitatory and one population of inhibitory neurons with strong lateral excitation and inhibition. Each excitatory cell in V4 projects to both excitatory and inhibitory neurons at its corresponding location in the saliency map, and receives feedback from excitatory neurons in this map (see supplemental Table 2, available at www.jneurosci.org as supplemental material, for model parameters). The saliency map also receives an input equal to 20% of the sum of the inputs to all feature-selective populations in V4 to account for direct inputs from the LGN and input layers of V1 to the LIP/FEF, which is supported by observation of an earlier response onset in the FEF with respect to V2 and V4 (Schmolesky et al., 1998). For simulation of the saccade preparation experiment, we use the same parameters while introducing extra inputs to all populations due to the presence of the saccade target (see below).

Inputs to the network. We use the same visual stimuli as used by Hegdé and Felleman (2003) and Burrows and Moore (2009) to generate the inputs to our model. These visual stimuli consist of arrays of 7 × 7 oriented, colored bars with six different arrangements: singleton (the single bar), homogeneous, color popout, orientation popout, combined popout, and conjunction (supplemental Fig. S1, available at www.jneurosci.org as supplemental material). Popout and conjunction displays contain one colored and oriented bar, the target, which can be distinguished from the rest of the colored and oriented bars, the 48 distractors, by either one or two features, respectively. In the above experiments, preferred and nonpreferred color and oriented bars were determined for each recorded neuron, and then used to construct different stimuli. For convenience, we construct the visual stimuli from four types of bars: green or red and vertical or horizontal. Moreover, we always place the target bar in the center of the array.

All neurons in the network receive two classes of synaptic inputs; background input and feature-selective inputs. The background input represents projections from surrounding cortical neurons and are modeled by Gaussian noise currents. For each excitatory (respectively inhibitory) neuron, this input is equal to the current generated by 1000 cortical neurons firing at 4.0 Hz (respectively 3.0 Hz), through AMPA synapses (with the peak conductance $g_s = 1$ nS). This spontaneous synaptic barrage provides the model with realistic noise and brings neurons near their firing threshold. The feature-selective inputs represent the outputs of color-selective neurons in the LGN, and of orientation-selective neurons in V1 (see below).

Responses to the orientation bars are computed using RF properties of orientation-selective cells in layer 4 of V1 (Dayan and Abbott, 2001).

More specifically, the input to location (x, y) at time t is equal to the following:

$$I_o(x,y,t) = I_o + kL_o(x,y,t), \quad (1)$$

where k is an arbitrary constant, and $L_o(x, y, t)$ is the linear response estimate of neuronal activity in space and time,

$$L_o(x,y,t) = \int_0^t dt' \int_{x_{min}}^{x_{max}} dx' \int_{y_{min}}^{y_{max}} dy' D_o(x',y',t')\, s(x-x', y-y'), \quad (2)$$

where $s(x, y)$ is the visual stimulus at location (x, y) and the kernel $D_o(x', y', t')$ defines the space-time RF of the neuron. As the input is stationary, s does not depend on time but is equal to the average intensity of that stimulus at a given location. The kernel can be decomposed into spatial and temporal RF components (Dayan and Abbott, 2001), as follows:

$$D_o(x',y',t') = D_{o,s}(x',y')\, D_{o,t}(t'). \quad (3)$$

The spatial RF of orientation-selective neurons is modeled with a Gabor function:

$$D_{o,s}(x',y') = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{x'^2}{2\sigma_x^2}-\frac{y'^2}{2\sigma_y^2}\right)\cos(kx'), \quad (4)$$

where σ_x and σ_y determine the extent of the RF, and $k = 1/8°$ is the preferred spatial frequency. In Equation 4, the neuron responds most strongly to 0° orientation. In our model, we employ four types of input neurons selective to different orientations (at 0°, 45°, 90°, 135°). We choose $\sigma_x = 0.5°$ for the nonpreferred direction, and $\sigma_y = 1°$ for the preferred direction. The temporal RF of the orientation-selective neurons is given by the following (Maex and Orban, 1996; Dayan and Abbott, 2001):

$$D_{o,t}(t') = \alpha\exp(-\alpha t')\left(\frac{(\alpha t')^5}{5!} - b_o\frac{(\alpha t')^7}{7!}\right), \quad (5)$$

where $\alpha = 1/(7.5\text{ ms})$ and $b_o = 0.85$. We introduce a bias factor, b_o, to make the integral of $D_{o,t}(t')$ nonzero, avoiding a vanishing response after the initial onset. Because the image input is stationary, the temporal component of the RF response can be integrated independently of the spatial component.

We assume that the color inputs to our network are generated by the response of the most prevalent type of color-selective neurons in the LGN (Ts'o and Gilbert, 1988). That is, the center region of RF has color-opponency while the surround RF is nonchromatic, matching the modified type II neurons in Ts'o and Gilbert (1988). We assume that the form of spatial and temporal RF of these neurons is similar to ON- and OFF-center neurons in the LGN with the kernel,

$$D_c(x',y',t') = \frac{D_{c,t}^{cen}(t')}{2\pi\sigma_{cen}^2}\exp\left(-\frac{x'^2+y'^2}{2\sigma_{cen}^2}\right) - \frac{BD_{c,t}^{sur}(t')}{2\pi\sigma_{sur}^2}\exp\left(-\frac{x'^2+y'^2}{2\sigma_{sur}^2}\right). \quad (6)$$

Here B is a constant which determines the relative contribution of surround to center, $\sigma_{cen} = 0.5°$ and $\sigma_{sur} = 1.0°$ determine the extent of RF in the center and surround, respectively, and $D_{c,t}^{cen}$ (respectively, $D_{c,t}^{sur}$) determines the temporal RF of the center (respectively, surround). We only use ON-center neurons to generate the inputs. The temporal component of the RF for the center and surround are described as follows (Dayan and Abbott, 2001):

$$D_{c,t}^{cen,sur}(t') = \alpha_{cen,sur}^2 t'\exp(-\alpha_{cen,sur}t') - b_c\beta_{cen,sur}^2 t'\exp(-\beta_{cen,sur}t'), \quad (7)$$

where $\alpha_{cen} = 1/(10\text{ ms})$, $\beta_{cen} = 1/(32\text{ ms})$, $\alpha_{sur} = 1/(20\text{ ms})$, $\beta_{sur} = 1/32$ ms), and $b_c = 0.75$ is a constant introduced to avoid the integral of the temporal RF vanishing over time. Because the input is stationary, the temporal component of the RF can be integrated independently of the spatial component.

To simulate red and green bars, the response of neurons with an RF described above is generated using three color components of the input image. The red response equals the center response to the red component minus the center response to green plus the center response to the average of the three components. The surround response is nonchromatic and is equal to the average of the three color components (Ts'o and Gilbert, 1988). This way, neurons selective to red show a strong response to a red bar in their RF, but a weak response to a green bar in their RF (they prefer red and then anything but green) and vice versa.

Simulated V2 and V4 neurons receive color- and/or orientation-selective inputs which are 30% and 15% of the selective inputs to V1 (described above), respectively. These inputs are implemented to account for weak direct inputs from the LGN and input layers of V1 to areas V2 and V4 (Girard and Bullier, 1989; Girard et al., 1991). Note that we obtain similar results even in the absence of these direct inputs to V2 and V4.

Finally, in the saccade preparation experiment of Burrows and Moore (2009) the monkey is cued to make a saccade to a target while visual stimuli are presented in the RF of recorded neurons at a random time before the saccade initiation. To simulate this experiment, we assume that the onset of the saccade target introduces a strong input to the saliency population and a weak input (equal to 20% of the input to the saliency population) to all feature-selective populations, at the location of the saccade target. The strong input to the saliency map is assumed to originate from a working memory network which encodes the location of the saccade target. For simplicity, we assume that the onset of the saccade target is always 50 ms before the onset of the visual stimuli.

Data analysis. For the results presented here, the average response of a simulated neuron to a given visual display is computed by counting all spikes in the 200 ms interval following the onset of activity (defined as firing above 5.0 Hz) and averaging this number over 200 trials. Because neighboring neurons have overlapping RF and their activity is highly correlated due to all-to-all connectivity, we compute these averages over four neurons with overlapping RF (for both target-selective or for distractor-selective neurons). For convenience, we present most of the average responses after normalizing them by the response to the singleton display.

To quantify the saliency computations in successive layers of the network, we compute different quantities. First, we consider the difference between the response to popout and conjunction displays. These differences can be used to define a local saliency measure known as the popout selectivity index (PI) which has been reported in many experimental studies (Knierim and van Essen, 1992; Hegdé and Felleman, 2003; Burrows and Moore, 2009):

$$PI = \frac{R_{popout} - R_{conj}}{R_{popout} + R_{conj}}, \quad (8)$$

where R_{popout} and R_{conj} are the average responses of target-selective neurons to a given popout and conjunction display, respectively.

Second, as the saliency of a target depends on its contrast with nearby objects, the neuronal signature of the saliency of a target should depend on the response of target-selective neurons relative to the response of neurons selective to nearby distractors. Therefore, we use the average response of neurons selective to the target and to all 48 distractors surrounding the target to define a global measure of saliency.

We define the global saliency index (GSI) as the difference in the average response of target-selective neurons (R_{target}) and of all distractor-selective neurons (R_{distr}) divided by their sum, as follows:

$$GSI = \frac{R_{target} - R_{distr}}{R_{target} + R_{distr}}. \quad (9)$$

GSI, distributed between -1 and $+1$, is a measure for how easily the target can be distinguished among the distractors: the closer it is to 1, the larger the neuronal representation of the target relative to the distractors.

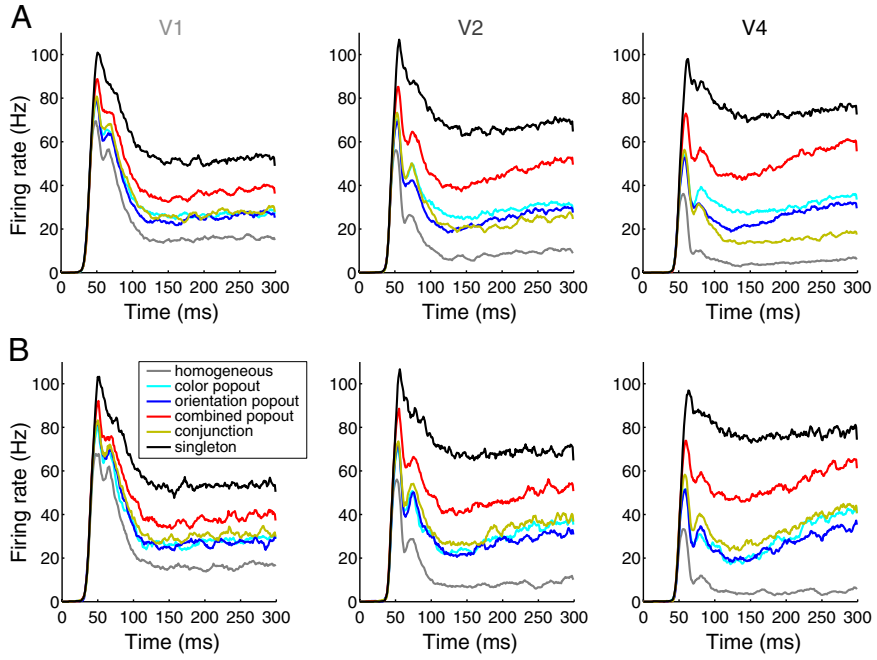

Figure 2. Simulated time course of the response of target-selective neurons in V1, V2 and V4 to the 6 displays. The responses are bounded from above by the response to the singleton (black curves) and from below by the response to the homogeneous array of bars (gray curves). ***A***, Configuration *A*, in which different features are processed independently. ***B***, Configuration *B*, in which combinations of features are jointly represented. The response in each condition is computed by averaging the response of four central neurons with overlapping RF for 200 trials of the simulated experiment. The response of target-selective neurons to the conjunction display falls below the response to popout displays in V4 for configuration *A* but not *B*.

Results

Saliency computations through lateral interactions in successive layers of spiking neurons

To understand the basic mechanisms underlying saliency computations, we examine two exclusive architectures (Fig. 1). In architecture or configuration *A*, individual features are processed in separate neural populations and neurons in different populations receive inputs selective to either orientation or color (Fig. 1*A*). We combine the outputs of such neural populations to construct a neural response selective to both features (see Materials and Methods for more details). In configuration *B*, a combination of features is jointly processed and neurons in different populations receive inputs selective to both orientation and color (Fig. 1*B*).

We first compare the response of neurons selective to the target (the central red-vertical bar) in V1, V2, and V4 and for the two configurations (Fig. 2). Note that the target is the same for all displays, while the surrounding distractors differ. The average response to the six displays are computed over 200 trials of network simulation. The onset of response (defined as firing above 5 Hz) to different stimuli occurs at ~40 ms for neurons in simulated V1, ~48 ms for V2, and ~54 ms for V4.

Soon after the activity onset, the responses to the different displays diverge due to lateral interactions in neural populations. Sometimes a smaller, secondary peak can be observed, a remnant of the shock oscillation caused by all-to-all connectivity. However, this oscillation is damped out quickly and does not contribute to the saliency computations. Because lateral interactions are dominated by inhibition, target-selective neurons show the weakest response to the homogeneous display, and simultaneously the strongest response to the singleton display (Fig. 2, compare black and gray traces). Note that when the singleton display is presented, the distractor-selective neurons receive the smallest inputs and so only weakly inhibit central neurons responding to the target. Yet when the homogeneous field of bars is presented, the distractor-selective neurons receive the largest inputs and so strongly inhibit target-selective neurons.

We also find that the response in V1 for both configurations for color and orientation popout displays (cyan and blue traces, respectively, in Fig. 2) are similar to the response to the conjunction display (yellow), while they are smaller than the response to the combined popout display (red). Thus, while target-selective V1 cells already show differential responses, they do not distinguish between popout and conjunction per se, similar to what has been observed in monkey V1 (see Hegdé and Felleman, 2003, their Fig. 10). As the signal propagates through V2 and V4, the response to the conjunction becomes smaller than the response to the orientation and color popout in configuration *A* but not in *B* (compare yellow traces in Fig. 2*A*,*B*). That is, neurons in higher areas show a differential response to the popout and conjunction displays only when individual features are processed independently.

To quantify the evolution of the saliency signal in successive regions, we next consider the average response to the 6 different displays in successive layers of the network. As expected, we find that the average response to all five arrays of bars are suppressed compared with the singleton for neurons selective to the target (supplemental Fig. S2, available at www.jneurosci.org as supplemental material). Interestingly, the response in V1 is fairly similar for both configurations *A* and *B* and qualitatively matches the experimental results in V1 (Hegdé and Felleman, 2003, their Fig. 5).

The saliency of a target depends on its contrast with nearby objects, here the neighboring distractors. Likewise, the neuronal signature of target saliency should depend on the response of target-selective neurons relative to the response of neurons selective to nearby distractors. Therefore, we further analyze the average response of both target- and distractor-selective neurons in each region and for each configuration (Fig. 3).

When individual features are processed separately, the response to the target in popout displays is reduced by small amounts from one region to the next, which is much smaller than the reduction in the homogeneous and conjunction displays (Fig. 3*A*). As a result, the average response to popout targets exceeds the response to the conjunction target for configuration *A*. This is accompanied by an increase in the difference between the response to the target and distractors in successive layers for popout displays, but not for the conjunction display. On the other hand, the response to the target in both popout and conjunction displays is reduced in configuration *B* which is accompanied by an increase in the difference between the response to the target and distractors in successive layers for these displays.(Fig. 3*B*). The differential response of target-selective neurons to popout and conjunction stimuli can be quantified by the popout selectivity index (Knierim and van Essen, 1992; Burrows and Moore, 2009). We find that V4 popout selectivity indices for config-

uration *A* are qualitatively similar to those measured for V4 neurons in the monkey (Fig. 3*C*; cf. Burrows and Moore, 2009, their Fig. 2*B*).

This differential response in the two configurations is a consequence of the fact that for configuration *B*, lateral interactions take place between neurons selective to combinations of features. Thus, the response of target-selective neurons to either color or orientation popout or to conjunction displays is suppressed by active distractor-selective neurons, while this is not the case in configuration *A*. In the latter, the distractor-selective neurons are active in only one of the two populations for color and orientation popout displays. Consequently, the response to popout and conjunction displays is suppressed differentially for configuration *A* but to the same extent for *B*. This effect is compounded when ascending through the three regions (Fig. 3), and results in activity patterns in V4 similar to experimental observations (Burrows and Moore, 2009), but only for configuration *A*. Therefore, we conclude that saliency computations require independent processing of individual features in successive layers of neurons with lateral interactions.

Although popout selectivity has been used as a measure of saliency signals, it has been argued that the absence of popout selectivity may not be equivalent to the absence of saliency signals (Li 2002). To test this hypothesis we use the average response of neurons selective to the target and all 48 distractors surrounding the target to define the global popout selectivity indices (see supplemental material, available at www.jneurosci.org). Interestingly, we find that local and global popout selectivity indices are highly correlated in all regions (see supplemental Figs. S4, S5, available at www.jneurosci.org as supplemental material). Local and global saliency signals are correlated because they are generated through lateral interactions between neurons with similar selectivity. Therefore, assuming saliency computation is manifested in the brain through the same mechanism, we conclude that both local and global popout indices can be equally informative about the visual saliency of the target.

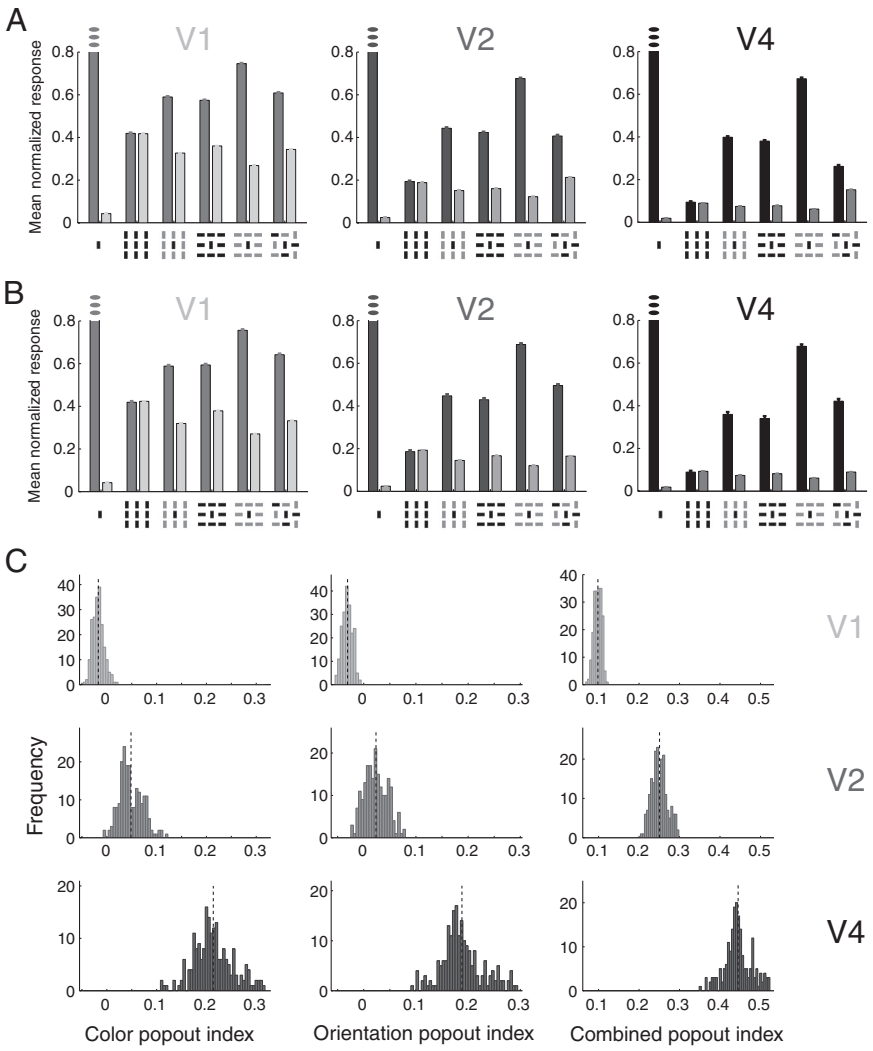

Figure 3. Evolution of response and the local saliency signal in feature-selective neurons in successive layers of the network. ***A***, ***B***, Average normalized response to different types of display is plotted, separately for neurons selective to either the target (central bar) or distractors (peripheral bars). In each panel, the average response of target-selective neurons is plotted with a darker shade than the average response of distractor-selective neurons. The error bars are the SEM. For illustrative purposes, we only show the values of responses between 0 and 0.8 (the response of target-selective neurons to the single bar is equal to 1). Results for configurations *A* and *B* are shown in ***A*** and ***B***, respectively. While in configuration *A* the response to the target decreases slightly, the difference between the response to the target and distractors increases only for popout displays in higher visual areas. In contrast, this difference increases for both popout and conjunction displays in configuration *B*. ***C***, Histograms of popout selectivity indices for popout displays in successive layers of the network. Histograms of popout indices are computed from the average response in the first 200 ms of the visual response in 30 randomly selected trials. Dashed lines show the mean value for each histogram.

Up to this point, we examined the response of target-selective neurons to different displays. We showed how a differential response to popout and conjunction displays is formed in successive layers of neurons through lateral excitatory and inhibitory interactions. The reason for examining the signal in target-selective neurons was to compare our results to electrophysiological recordings, although the presence of this signal is not equivalent to target detection per se (see Discussion). Instead, target selection can be performed by finding the most active location in a topographic map which is not selective to any feature (Itti and Koch, 2001). Therefore, we construct a "hypothetical" saliency map by adding the output of feature-selective neurons in each region to examine the signal related to target detection.

We find that the fictitious saliency neurons in early visual areas show little or no difference in response to the target and distractors. However, for the network architecture *A*, differential responses emerge in higher visual areas for the singleton and popout displays but not for the conjunction display (Fig. 4*A*). More specifically, the global differential response (i.e., the difference between the response of target- and distractor-selective neurons) increases for popout, while it fluctuates around zero for the conjunction display in all three regions (Fig. 4*B*). This is not the case for the network architecture *B*, where the global differential response also increases for the conjunction display (supplemental Figs. S6, S7, available at www.jneurosci.org as supplemental material).

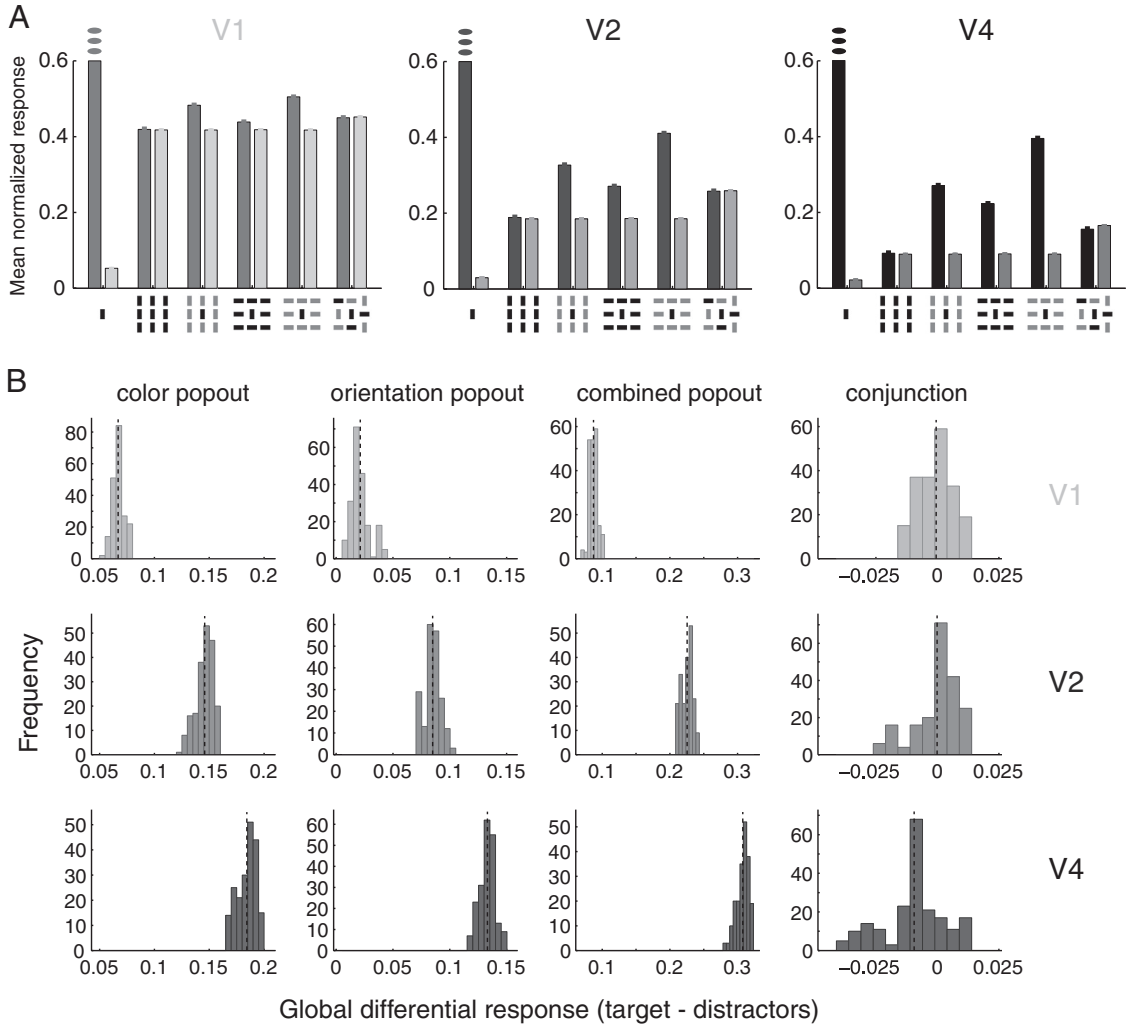

Figure 4. Evolution of the global saliency signal in the constructed saliency populations in successive regions for configuration *A*. ***A***, Average normalized response of the constructed saliency populations to different displays is plotted separately for neurons selective to either the target (central bar) or distractors (peripheral bars). The response of target-selective neurons is plotted with a darker shade than the response of distractor-selective neurons. The error bars are the SEM. For illustrative purposes, we only show the values of responses between 0 and 0.6 (the response to the single bar is equal to 1). ***B***, Histograms of the global differential response (i.e., the difference between response of target- and all distractor-selective neurons), for different displays. Histograms of the differential response are computed from the average response in the first 200 ms of the visual response in 30 randomly selected trials. The dashed lines show the mean in each histogram. The differential response increases for popout displays in successive regions but slightly decreases for the conjunction display.

Compatible with these observations, we find that when individual features are processed in distinct populations of neurons, the detection of popout but not conjunction targets is improved in successive layers in configuration *A* but not *B* (Fig. 5). Overall, these results indicate that a feature-independent saliency signal can be formed by convergence of outputs of different feature-selective neural populations in higher areas of the visual cortex, but this mechanism is effective only if feature processing is kept separated in lower visual areas.

When do local and global saliency signals first emerge? We examined the time course of the local saliency signal by calculating the difference between the activity of target-selective neurons in response to a given popout and conjunction displays: not only does this difference increase but it emerges earlier relative to the activity onset, as signals propagate from V1 to V2 and V4 (Fig. 6*A*).

We likewise examined the time course of the global saliency signal by calculating the difference between the response of target-selective neurons and of the most active distractor-selective neurons on each trial (for neurons in the constructed saliency populations). This difference increases and occurs earlier relative to the response onset in higher regions, but only for popout displays (Fig. 6*B*). Note that at the beginning of a trial before saliency computations are formed through lateral interactions, neurons are mainly driven by external inputs and because of noise many neurons can have higher activity than target-selective neurons. This results in early negative values for global saliency signals in our model.

Role of NMDA in saliency computations

The excitatory currents in our network model is transmitted through two types of synapses, fast AMPA and slow saturating NMDA. Generally, we find that the recurrent excitation should be dominated by NMDA currents but to test the role of these synapses in saliency computations, we reduce NMDA currents while increasing AMPA currents such that the overall recurrent excitation stays at the same level (see Materials and Methods for more details).

We find that increasing the AMPA to NMDA current ratio disrupts the formation of both local and global saliency signals (Fig. 7). That is, the response to popout displays is not different

from the response to the conjunction display in higher visual areas (Fig. 7*A*), and similarly, the difference in response of target- and distractor-selective neurons in the constructed saliency populations is reduced (Fig. 7*B*).

Moreover, as the NMDA currents are reduced, the amount of increase in the local and global differential response in successive regions is also reduced. Similarly we find that the probability of maximum response at the target location is reduced, especially for color and orientation popout displays (supplemental Fig. S8, available at www.jneurosci.org as supplemental material). Finally, we compute the temporal dynamics of the saliency signal for different values of the AMPA to NMDA current ratio. We find that an increase in the AMPA to NMDA current ratio delays and further eliminates the formation of the saliency signal in higher visual areas and moreover, introduces a strong oscillation in the response in these regions (supplemental Figs. S9, S10, available at www.jneurosci.org as supplemental material).

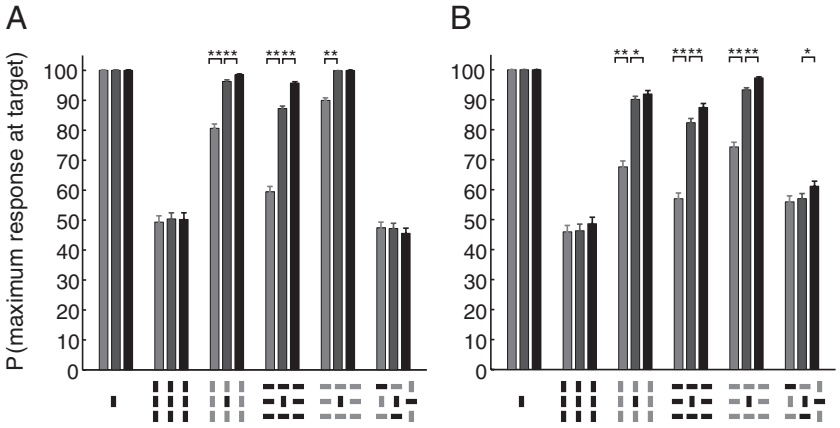

Figure 5. Improvement in target detection in successive layers of the network. The probability that the response of target-selective neurons is the maximum response (*P*(maximum response at target)) in the constructed saliency population is plotted with different shades of gray corresponding to three regions for different displays: light gray (V1), medium gray (V2), and black (V4). This probability is computed as the fraction of distractor-selective neurons which show smaller response than the target-selective neurons on each trial. The error bars are the SEM. One (respectively two) asterisk shows the statistical test that the described probability increases from one region to the next is significant at $p < 0.01$ (respectively, $p < 0.001$). Results for configurations *A* and *B* are shown in ***A*** and ***B***, respectively. In configuration *A*, *P*(maximum response at target) for popout displays reaches to 100% in the third layer, while this quantity is not different from 50% for the conjunction display. On the other hand for configuration *B*, *P*(maximum response at target) does not reach to 100% for color and orientation popout displays and furthermore, it increases in higher visual areas for the conjunction display.

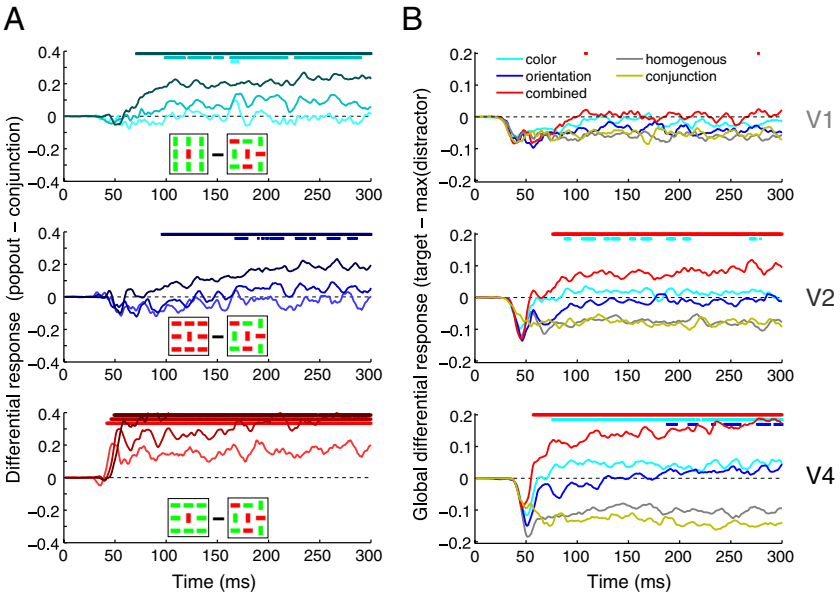

Figure 6. Temporal dynamics of local and global saliency signals. ***A***, The difference in the normalized response to the popout and conjunction displays is plotted for different types of popout displays (indicated by insets). Different shades of color (light to dark) correspond to responses in successive regions (V1 to V4, respectively). ***B***, The difference in the normalized response of the constructed saliency map neurons selective to the target and of neurons selective to the distractor that exhibit the maximum activity on a given trial is plotted for different types of displays and in three regions. The point at the top of each panel shows whether the differential response is statistically significant (at $p < 0.01$) for each 10 ms time interval. Both local and global saliency signals are larger in higher visual areas and emerge earlier in these areas relative to the response onset in these areas.

An explicit saliency map and its action onto earlier stages

To study the effect of feedback from the saliency map on the saliency computations in lower visual areas, we use a smaller version of our model with only two layers of neural populations corresponding to neurons in V4 and LIP/FEF (due to computational expenses). As we show in the previous sections, the saliency signal can be formed in successive layers of neurons when individual features are processed separately. Therefore, here we use the same architecture for neurons in V4 (see Materials and Methods for more details).

We first assure ourselves that approximate saliency signals can form in a single simulated cortical area (compared with three as in the above simulations). We use a stronger and more widely projecting regional connectivity matrix than before (compare model parameters in supplemental Tables 1, 2, available at www.jneurosci.org as supplemental material) to reproduce our basic result: the response of color/orientation-selective neurons to all popout displays is larger than the response to the conjunction display (Fig. 8*A*). Nevertheless, we observe a small positive but significant global saliency index for the conjunction display, which does not appear in saliency computations in successive regions of V1, V2 and V4 (compare Fig. 8 and results for V4 in Fig. 4*B*).

Because the input to the saliency map is the sum of the outputs of feature-selective neurons, this input carries a smaller saliency signal than the population of neurons selective to the target. That is, the global differential response is larger in the V4 population selective to red-vertical bars than in the saliency map (compare the response to the target and distractors for each case in Fig. 8*A*). We will return to this issue in the Discussion.

To examine the result of lateral interactions in the saliency populations, we then compute *GSI* for the synaptic inputs that originate from V4 cells (to the saliency map) and for the response

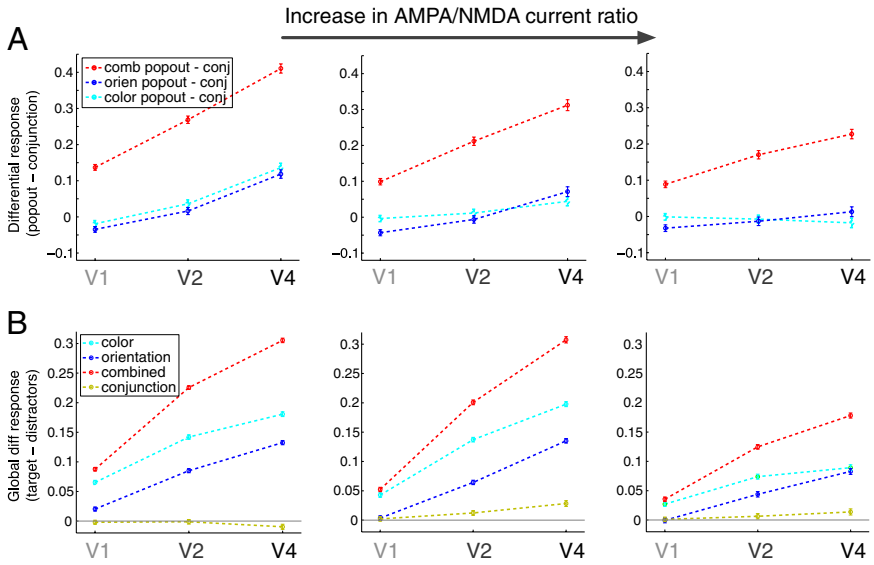

Figure 7. Local and global saliency signal formations are disrupted by increasing the AMPA to NMDA current ratio. ***A***, The difference between average response to different popout displays and to the conjunction display, for different AMPA to NMDA current ratios. The error bars are the SEM. ***B***, The average global differential response in the constructed saliency populations (i.e., difference between response of target- and distractor-selective neurons), for different types of displays and in successive regions of the network. As the AMPA to NMDA current ratio increases, the enhancement of both local and global saliency signals in successive regions is reduced or even diminished.

of neurons in the saliency map. As expected, we find that the lateral interactions within the saliency map enhance the *GSI* and therefore enhance the saliency signal (Fig. 8*B*). Note that the observed decrease in the normalized response in the saliency map population (compared with its inputs) is due to the large response to the singleton display in this population.

We next introduce feedback from the excitatory neurons in the saliency map to all excitatory and inhibitory populations in V4 at corresponding locations. These two types of feedback are adjusted such that they approximately modulate the response of the feature-selective neurons rather than driving them (Schwabe et al., 2006). Feedback differentially changes the response to each display, increasing popout selectivity while increasing variability (Fig. 9).

Finally, how does this model act in the presence of a top-down signal, as in the saccade preparation experiment of Burrows and Moore (2009)? In this experiment, the monkey was cued to make a saccade to a location far from the RF of the recorded V4 neuron, and was rewarded a drop of juice for making saccade to the cued location after the fixation spot disappeared. While planning such a saccade, a visual stimulus was presented at a random time before the initiation of the saccade. They found that under this manipulation, the observed differential response of V4 neurons to the popout and conjunction displays in the control passive viewing experiment, was eliminated. We hypothesize that interaction between bottom-up and top-down attentional signals occurs in the saliency map (LIP or FEF), where both signals are found (Thompson et al., 2005; Balan and Gottlieb, 2006; Ipata et al., 2006; Buschman and Miller, 2007). More specifically, we assume that saccade preparation induces an activity within the saliency map at the saccade target location. This corresponds to the introduction a highly salient target at that location through working memory, thereby altering feedback from the saliency map to earlier areas.

To test this hypothesis, we simulate the saccade preparation task by adding excitatory inputs to all populations at locations corresponding to the saccade target (to mimic working memory inputs). These inputs are strong enough so there is a representation of the saccade target in the saliency map on all trials, but are weak enough so they do not alter the response to the singleton display.

During the simulated saccade preparation the response to different displays changes slightly and differentially (Fig. 9), such that the popout selectivity indices for popout are reduced, in line with the experimental data (Burrows and Moore, 2009). That is, saccade preparation reduces the response to popout displays more so than the response to the conjunction display. This happens because for popout displays, the target provokes a strong response in the saliency map (Fig. 8*A*) which in turn results in a strong feedback to feature-selective neurons. The converse is true for the conjunction target. During saccade preparation, the saccade target also invokes a strong response in the saliency population which reduces the response to the target in this population through lateral inhibition (supplemental Fig. S11, available at www.jneurosci.org as supplemental material) and consequently, the response to the target in feature-selective neurons. Therefore, the amount of reduction in the response of feature-selective neurons depends on the target response in the saliency map during the control trials, and on the response to the saccade target in this map.

Discussion

Despite a large body of literature on the psychology of bottom-up attention and how it operates within visual scenes, much less is known about its neural substrates. Here we design a biophysically plausible spiking network model to investigate the representation and formation of saliency signals in the visual cortex and its interaction with top-down attention.

We focus on lateral excitatory and inhibitory interactions as the main mechanism for saliency computations. By comparing two distinct network architectures, we find that local and global saliency signals emerge and increase in successive layers of neural populations only if individual features are processed in different neural populations (configuration *A*). That is, while the activity of target-selective neurons in the first visual area of our model (V1) does not discriminate between popout and conjunction displays, neurons in higher areas of the model (V2 and V4) show stronger response to popout than to conjunction displays, similar to experimental observations in V1 and V4, respectively (Hegdé and Felleman, 2003; Burrows and Moore, 2009). Moreover, the difference between the response to the target and distractors, as well as target detectability, increases in successive layers for popout but not for conjunction displays, compatible with the basic difference in detection of popout and conjunction targets (Treisman and Sato, 1990).

Similar to experimental data (Burrows and Moore, 2009) we obtain larger local and global saliency signals for the combined popout than for single popout displays; that is, the detection of the popout target is easier when it differs from distractor in two features rather than one feature. Our finding is also consistent with the so-called "redundant-signal effect" (shortening of the reaction time when the response is triggered by two rather than

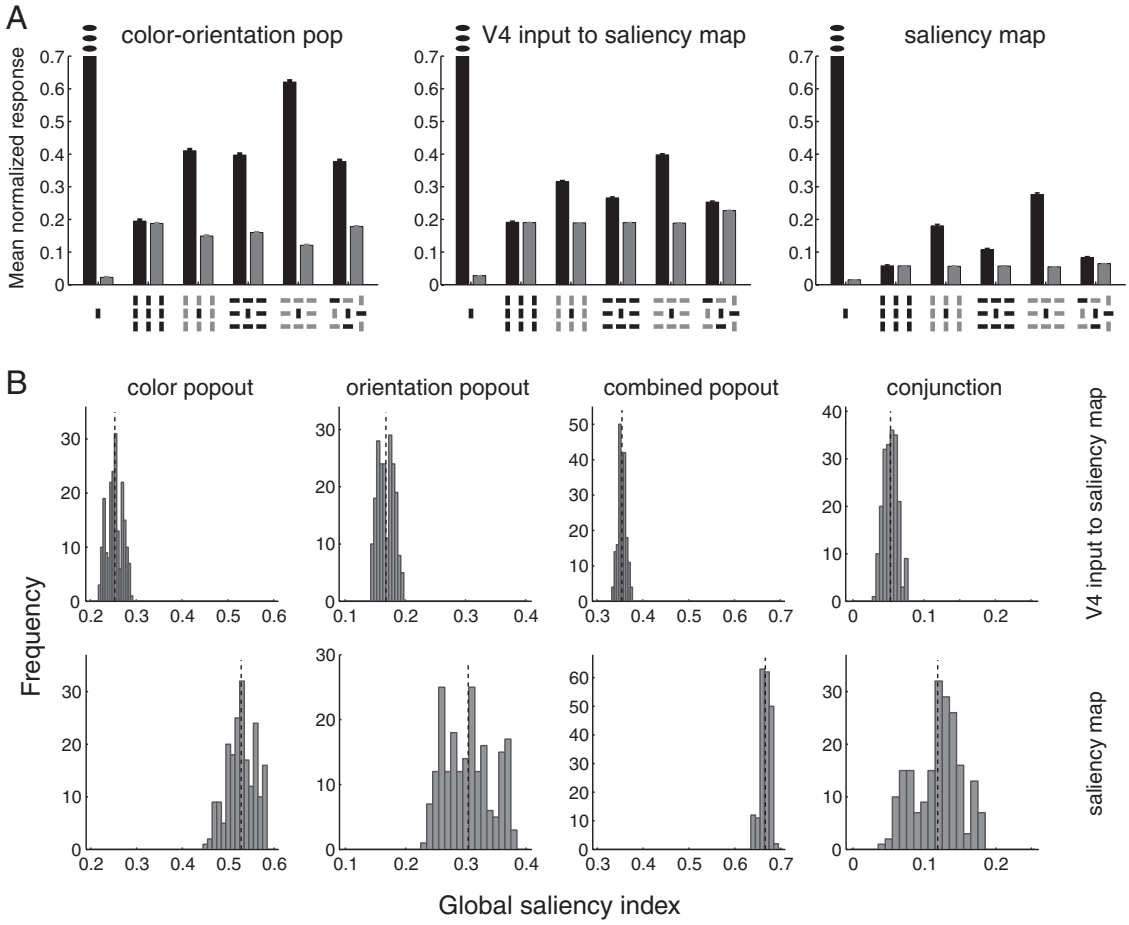

Figure 8. Formation of the saliency signal in a single, simulated cortical region V4 with stronger and wider lateral interactions, and in the saliency map which receives the output of these V4 neurons. ***A***, Average normalized response to different displays is plotted separately for neurons selective to either the target (black) or distractors (gray) in different populations: color/orientation-selective population, input to the saliency map from V4, and the saliency map population. The measured color-orientation neurons are selective to red-vertical bars, and the error bars show the SEM. For illustrative purposes, we only show the values of responses between 0 and 0.7 (the response to the single bar is equal to 1). ***B***, Histograms of the global saliency indices (*GSI*) for different displays are plotted for the input from V4 into the saliency map and for neurons in the saliency map. Dashed lines show the mean of the global saliency index for each display. The effect of lateral interactions in the saliency map is to increase the mean and variance of the global saliency indices.

one response-related target signal) demonstrated in a visual popout search task. This effect has been attributed to coactivation of different visual pathways and their subsequent convergence before response triggering (Krummenacher et al., 2001, 2002, 2010; Zehetleitner et al., 2008; Töllner et al., 2010). Similarly, in our model saliency signals for the combined popout is stronger because of independent input processing related to different features and their parallel processing in separate neural populations before convergence in the saliency map.

As an alternative to saliency computations in successive layers, we also consider wider and stronger lateral interactions in one layer of neural populations which process individual features separately. Even though we observe local saliency signals (i.e., positive popout selectivity indices), this signal was very small for orientation popout display (Fig. 9*B*). Moreover, this mechanism results in a small positive but significant global saliency index for the conjunction display (Fig. 8) and in a noisier detection of the target (data not shown).

Therefore, we conclude that moderate lateral interactions in successive layers of neurons selective to individual features provide a suitable mechanism for early saliency computations. Furthermore, neurons that process individual features separately are more likely to contribute to bottom-up saliency than neurons that are simultaneously selective to both color and orientation (Livingstone and Hubel, 1987).

Our saliency computations can be compared with the standard Itti-Koch computational saliency model (Koch and Ullman, 1985; Itti et al., 1998; Itti and Koch, 2001). The saliency model exploits center-surround computations (i.e., subtracting a filtered image at a lower spatial resolution from the image at a higher spatial resolution) to capture local feature contrasts in the image and to form feature maps, as well as normalization to enhance (respectively suppress) those maps with a few (respectively many) active locations. Lateral excitation and inhibition between neural populations enables our model to approximately perform center-surround and normalization computations without using a multi-resolution representation of an input image at different scales. An important requirement for this to happen is that inhibitory connections should be wider than excitatory connections.

Normalization of sensory inputs by the sum of inputs has been used in a few models of top-down attention (Reynolds et al., 1999; Lee and Maunsell, 2009; Reynolds and Heeger, 2009). For example, Reynolds and Heeger (2009) proposed that top-down attention improves sensitivity to faint stimulus through multiplicative interaction of inputs and the "attention field," followed by

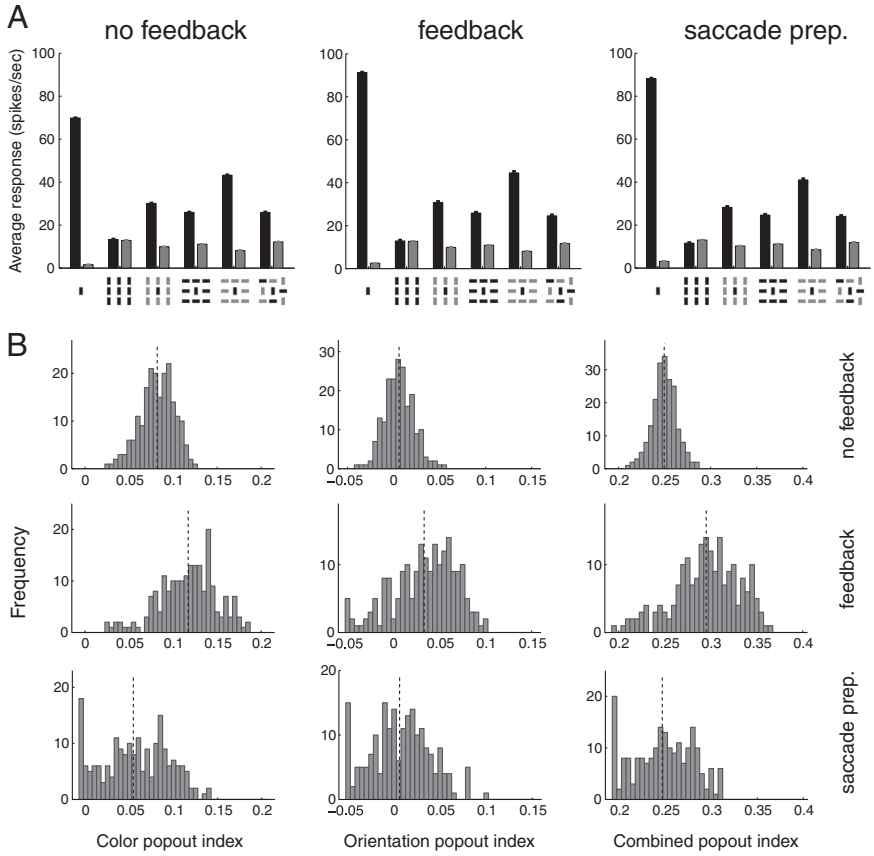

Figure 9. The effects of feedback from the saliency map and saccade preparation on saliency computations. ***A***, Average response to different displays is plotted separately for neurons selective to either the target (black) or distractors (gray) in three different cases: with no feedback, with feedback from the saliency map to all feature-selective populations, and simulated saccade preparation experiment. The measured neurons are selective to red-vertical bars, and the error bars show the SEM. ***B***, Histograms of the popout selectivity index for three different types of popout in three cases described above. The dashed lines show the mean of the popout index for each histogram. All popout selectivity indices are increased due to feedback ($p < 0.0001$) and are then decreased during the saccade preparation ($p < 0.0001$).

normalization of the response by the activity in the "suppressive field." Lateral interactions proposed here can provide a biophysical mechanism for normalization due to the suppressive field in this model. Alternatively, Lee and Maunsell (2009) proposed that top-down attention affects neural response solely through changing the strength of normalization and not the inputs. If such a mechanism is implemented through lateral interactions, then top-down attention should mostly affect the activity in inhibitory neurons to change the strength of normalization.

While lateral interactions between spiking neurons through realistic synapses can approximate center-surround and normalization computations, these computations are limited by biophysical properties of neural elements in the network. This happens because neurons in the network should integrate, in every spike, the noisy background inputs plus excitatory and inhibitory inputs from neighboring neurons and subsequently, transmit this signal to other neurons in the network through realistic synapses with different time constants. Due to these factors, some conditions should be met in order for the network to perform efficient center-surround and normalization computations. First, we find that recurrent excitations between excitatory neurons should be dominated by slow NMDA currents. This enables neurons to integrate the noisy input signals over a longer timescale and to increase the signal-to-noise ratio. Second, excitatory to inhibitory connections, which drive the lateral inhibition through inhibitory interneurons, should have both AMPA and NMDA currents. This is because if these connections are also dominated by NMDA currents, the inhibitory interneurons become active slowly which can suppress the activity in the network after the response onset. These results emphasize the crucial role of the NMDA receptor in the saliency computations, similar to its role in working memory and decision making (Wang, 1999, 2002; Compte et al., 2000). Based on our results, we predict that inactivation of NMDA receptors in the visual cortex results in a nosier and weaker saliency signal, and impairs the performance in the visual search task.

Although we find that recurrent excitation in our spiking network should be dominated by slow NMDA currents, this does not translate to slow emergence of the saliency signal. Interestingly, we find that both local and global saliency signals emerge earlier in successive regions of the model (relative to the response onset in each region). Similarly, Buffalo and colleagues recently showed that top-down attentional effects appear earlier and stronger in V4 than in V2 and V1 (Buffalo et al., 2010). This result may also explain how the saliency signal can appear in higher visual areas such as the LIP and FEF even before the appearance of this signal in striate and extrastriate cortices. For example, compare the emergence of this signal in V4 (Burrows and Moore, 2009) with the LIP and FEF (Buschman and Miller, 2007). Interestingly, Buschman and Miller (2007) found an earlier saliency signal in the LIP than in the FEF during popout search, while this signal emerges earlier in the FEF during a visual search task which requires top-down attention (but see Schall et al., 2007). These observations suggest different roles for the LIP and FEF in saliency computations, a topic which requires further investigations.

V1 has been implicated as the site of early saliency signals (Li, 2002) because V1 neurons are influenced by the stimulation of regions outside their classical RF in a nonlinear fashion (DeAngelis et al., 1994; Albright and Stoner, 2002; Cavanaugh et al., 2002a,b). For example, activity of V1 neurons in the alert monkey is only weakly suppressed (with respect to the singleton display) when the surrounding bars are in orientation perpendicular to the orientation of the central bar (Knierim and van Essen, 1992). However, an experiment specifically designed to detect the presence of saliency signals in monkey V1 came up empty handed; Hegdé and Felleman (2003) found that V1 neurons do not distinguish between popout and conjunction displays; rather, they signal the existence of center-surround discontinuity. Similarly, we find that moderate lateral interactions between neurons with similar feature selectivity can result in a response pattern which depends on the display but does not distinguish between popout and conjunction. In addition, we show that the absence of local saliency signals is indicative of the absence of global saliency signals.

Conversely, single cell data indicates that V4 neurons are selective to bottom-up attentional signals such as popout display, and are modulated by top-down attention as well as the activity of neurons in LIP and FEF (Schiller and Lee, 1991; Hupé et al., 1998; Reynolds et al., 2000; Tolias et al., 2001; Moore and Armstrong, 2003; Reynolds and Desimone, 2003; Armstrong et al., 2006; Armstrong and Moore, 2007; Gregoriou et al., 2009). Compelling evidence for the presence of saliency signals in V4 is found by Burrows and Moore (2009). They demonstrated that V4 neurons, considered as a single population, respond stronger to popout than to conjunction displays. Furthermore, this difference is eliminated when the monkey prepares a saccade to a location far from the RF of the recorded neuron, which indicates that this bottom-up saliency signal is influenced by top-down attentional signals. There is converging evidence that these signals possibly originate in the FEF (Moore and Armstrong, 2003; Armstrong et al., 2006; Armstrong and Moore, 2007; Monosov et al., 2008; Gregoriou et al., 2009). We find in our model that feedback from the saliency map to earlier regions enhances the saliency signal while saccade preparation reduces this signal by altering the feedback.

Interestingly, our model predicts that the effect of saccade preparation on the response to a given stimulus depends on the response of target-selective neurons in the saliency map during the control passive viewing task. This prediction can be tested by recording from neurons in the saliency map (e.g., LIP/FEF) and in feature-selective populations which receive feedback from the saliency map (e.g., V4). Such recording can be used to compute the correlation between the responses of neurons in the saliency map and the reduction in the popout selectivity index for different displays in the control and saccade preparation tasks, respectively. We predict a positive correlation between these two quantities.

By combining the outputs of neural populations which process individual features (in configuration *A*) we construct different color/orientation-selective neural populations. Among these populations the one which is selective to the target features carries a saliency signal stronger than the signal in the saliency population (Fig. 8). So why should the brain bother with a distinct saliency map in the first place? However, in the absence of an explicit saliency map, the brain needs to detect the target by first identifying this feature-selective population. This is of course quite difficult in dense natural scenes with many, partially occluded targets, which is why the strategy of a saliency map that labels the sites of potential objects is an attractive computational option. Similarly, feedback to neural populations in V1 from V2 or V4 populations with similar selectivity does not improve saliency signals as this feedback does not contain any information about the most salient location in the visual scene, and only acts as a stronger recurrent input from the same area. Instead, feedback from the saliency map improves saliency signal as it can enhance the signal in neurons selective to the most salient location. Finally, a feature-independent saliency map, formed by the convergence of outputs of neural populations selective to different features, is consistent with the observation that the saliency signal in the FEF appears in the spiking activity before the local field potential (Monosov et al., 2008).

Electrophysiological evidence suggests that a saliency map is instantiated in the response of neurons in the posterior parietal area 7a (Constantinidis and Steinmetz, 2001, 2005), area LIP (Gottliebet al., 1998; Kusunoki et al., 2000; Bisley and Goldberg, 2003), and in the FEF (Thompson and Bichot, 2005). Interestingly, in such a setting the activity of the saliency population and also the detection of the target can be adjusted by gating the inputs from different feature-selective populations (Rutishauser and Koch, 2007) and by top-down signals. To model saccade preparation, we assume that saliency neurons selective to the location of the saccade target become active and stay active during the stimulus presentation (at a fixed level of activity). Conceivably, trial-by-trial variability in the representation of the saccade target can alter the feedback to V4 neurons and consequently, the saliency signal. Such variability has been observed in areas LIP and FEF and was shown to be correlated with monkeys ability to ignore distractors (Thompson et al., 2005; Balan and Gottlieb, 2006; Ipata et al., 2006).

At the end, while it has been shown that saliency computations can be easily performed through center-surround and normalization computations (Itti et al., 1998), we find biophysical limits for performing these computations by spiking neurons and realistic synapses. These limits point to the general biophysical mechanisms which are used in other parts of the brain. Namely, due to response variability of cortical neurons, the integration of input signals need to be done through slow NMDA synapses and in successive layers of neural populations. Computation in successive layers of neural populations results in earlier emergence of the saliency signal in higher visual areas, which in turn can provide feedback to lower visual areas and improve the saliency computations.

References

Albright TD, Stoner GR (2002) Contextual influences on visual processing. Annu Rev Neurosci 25:339–379.

Allman J, Miezin F, McGuinness E (1985) Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. Annu Rev Neurosci 8:407–430.

Armstrong KM, Moore T (2007) Rapid enhancement of visual cortical response discriminability by microstimulation of the frontal eye field. Proc Natl Acad Sci U S A 104:9499–9504.

Armstrong KM, Fitzgerald JK, Moore T (2006) Changes in visual receptive fields with microstimulation of frontal cortex. Neuron 50:791–798.

Balan PF, Gottlieb J (2006) Integration of exogenous input into a dynamic salience map revealed by perturbing attention. J Neurosci 26:9239–9249.

Bisley JW, Goldberg ME (2003) Neuronal activity in the lateral intra parietal area and spatial attention. Science 299:81–86.

Buffalo EA, Fries P, Landman R, Liang H, Desimone R (2010) A backward progression of attentional effects in the ventral stream. Proc Natl Acad Sci U S A 107:361–365.

Burrows BE, Moore T (2009) Influence and limitations of popout in the selection of salient visual stimuli by area V4 neurons. J Neurosci 29: 15169–15177.

Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315:1860–1862.

Cavanaugh JR, Bair W, Movshon JA (2002a) Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. J Neurophysiol 88:2530–2546.

Cavanaugh JR, Bair W, Movshon JA (2002b) Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. J Neurophysiol 88:2547–2556.

Cerf M, Harel J, Einhäuser W, Koch C (2008) Predicting human gaze using low-level saliency combined with face detection. Adv Neur Inf Proc Sys 20:241–248.

Compte A, Brunel N, Goldman-Rakic PS, Wang XJ (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cereb Cortex 10:910–923.

Constantinidis C, Steinmetz MA (2001) Neuronal responses in area 7a to multiple-stimulus displays: I. Neurons encode the location of the salient stimulus. Cereb Cortex 11:581–591.

Constantinidis C, Steinmetz MA (2005) Posterior parietal cortex automatically encodes the location of salient stimuli. J Neurosci 25:233–238.

Dayan P, Abbott LF (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge, MA: MIT.

DeAngelis GC, Freeman RD, Ohzawa I (1994) Length and width tuning of neurons in the cat's primary visual cortex. J Neurophysiol 71:347–374.
Foulsham T, Underwood G (2008) What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. J Vis 8:6.1–6.17.
Girard P, Bullier J (1989) Visual activity in area V2 during reversible inactivation of area 17 in the macaque monkey. J Neurophysiol 62:1287–1302.
Girard P, Salin PA, Bullier J (1991) Visual activity in macaque area V4 depends on area 17 input. Neuroreport 2:81–84.
Gottlieb JP, Kusunoki M, Goldberg ME (1998) The representation of visual salience in monkey parietal cortex. Nature 391:481–484.
Gregoriou GG, Gotts SJ, Zhou H, Desimone R (2009) High-frequency, long-range coupling between prefrontal and visual cortex during attention. Science 324:1207–1210.
Hansel D, Mato G, Meunier C, Neltner L (1998) On numerical simulations of integrate-and-fire neural networks. Neural Comp 10:467–483.
Hegdé J, Felleman DJ (2003) How selective are V1 cells for pop-out stimuli? J Neurosci 23:9968–9980.
Hupé JM, James AC, Payne BR, Lomber SG, Girard P, Bullier J (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 394:784–787.
Ipata AE, Gee AL, Gottlieb J, Bisley JW, Goldberg ME (2006) LIP responses to a popout stimulus are reduced if it is overtly ignored. Nat Neurosci 9:1071–1076.
Itti L, Koch C (2001) Computational modelling of visual attention. Nat Rev Neurosci 2:194–203.
Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Patt Anal Mach Intell 20:1254–1259.
Knierim JJ, van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J Neurophysiol 67:961–980.
Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol 4:219–227.
Krummenacher J, Müller HJ, Heller D (2001) Visual search for dimensionally redundant pop-out targets: evidence for parallel-coactive processing of dimensions. Percept Psychophys 63:901–917.
Krummenacher J, Müller HJ, Heller D (2002) Visual search for dimensionally redundant pop-out targets: parallel-coactive processing of dimensions is location specific. J Exp Psychol Hum Percept Perform 28: 1303–1322.
Krummenacher J, Grubert A, Müller HJ (2010) Inter-trial and redundant-signals effects in visual search and discrimination tasks: separable pre-attentive and post-selective effects. Vision Res 50:1382–1395.
Kusunoki M, Gottlieb J, Goldberg ME (2000) The lateral intraparietal area as a salience map: the representation of abrupt onset, stimulus motion, and task relevance. Vision Res 40:1459–1468.
Lee J, Maunsell JH (2009) A normalization model of attentional modulation of single unit responses. PLoS One 4:e4651.
Li Z (2002) A saliency map in primary visual cortex. Trends Cogn Sci 6:9–16.
Livingstone MS, Hubel DH (1987) Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. J Neurosci 7:3416–3468.
Maex R, Orban GA (1996) Model circuit of spiking neurons generating directional selectivity in simple cells. J Neurophysiol 75:1515–1545.
Mannan SK, Kennard C, Husain M (2009) The role of visual salience in directing eye movements in visual object agnosia. Cur Biol 19:R247–R248.
Monosov IE, Trageser JC, Thompson KG (2008) Measurements of simultaneously recorded spiking activity and local field potentials suggest that spatial selection emerges in the frontal eye field. Neuron 57:614–625.
Moore T, Armstrong KM (2003) Selective gating of visual signals by microstimulation of frontal cortex. Nature 421:370–373.
Parkhurst D, Law K, Niebur E (2002) Modeling the role of salience in the allocation of overt visual attention. Vision Res 42:107–123.
Reynolds JH, Desimone R (2003) Interacting roles of attention and visual salience in V4. Neuron 37:853–863.
Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron: 61:168–185.
Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas v2 and v4. J Neurosci 19:1736–1753.
Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. Neuron 26:703–714.
Rutishauser U, Koch C (2007) Probabilistic modeling of eye movement data during conjunction search via feature-based attention. J Vis 7:1–20.
Schall JD, Paré M, Woodman GF (2007) Comment on 'Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices'. Science 318:44 [author reply 44].
Schiller PH, Lee K (1991) The role of the primate extrastriate area V4 in vision. Science 251:1251–1253.
Schmolesky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AG (1998) Signal timing across the macaque visual system. J Neurophysiol 79:3272–3278.
Schwabe L, Obermayer K, Angelucci A, Bressloff PC (2006) The role of feedback in shaping the extra-classical receptive field of cortical neurons: a recurrent network model. J Neurosci 26:9117–9129.
Thompson KG, Bichot NP (2005) A visual salience map in the primate frontal eye field. Prog Brain Res 147:251–262.
Thompson KG, Bichot NP, Sato TR (2005) Frontal eye field activity before visual search errors reveals the integration of bottom-up and top-down salience. J Neurophysiol 93:337–351.
Tolias AS, Moore T, Smirnakis SM, Tehovnik EJ, Siapas AG, Schiller PH (2001) Eye movements modulate visual receptive fields of V4 neurons. Neuron 29:757–767.
Töllner T, Zehetleitner M, Krummenacher J, Müller HJ (2010) Perceptual basis of redundancy gains in visual pop-out search. J Cogn Neurosci. Advance online publication. Retrieved May 25, 2010. doi:10.1162/jocn.2010.21422.
Treisman A, Sato S (1990) Conjunction search revisited. J Exp Psychol Hum Percept Perform 16:459–478.
Ts'o DY, Gilbert CD (1988) The organization of chromatic and spatial interactions in the primate striate cortex. J Neurosci 8:1712–1727.
Wang XJ (1999) Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. J Neurosci 19:9587–9603.
Wang XJ (2002) Probabilistic decision making by slow reverberation in cortical circuits. Neuron 36:955–968.
Zehetleitner M, Muller HJ, Krummenacher J (2008) The redundant-signals paradigm and preattentive visual processing. Front Biosci 13:5279–5293.

Behavioral/Systems/Cognitive

A Biophysically Based Neural Model of Matching Law Behavior: Melioration by Stochastic Synapses

Alireza Soltani and Xiao-Jing Wang
Volen Center for Complex Systems and Department of Physics, Brandeis University, Waltham, Massachusetts 02454

In experiments designed to uncover the neural basis of adaptive decision making in a foraging environment, neuroscientists have reported single-cell activities in the lateral intraparietal cortex (LIP) that are correlated with choice options and their subjective values. To investigate the underlying synaptic mechanism, we considered a spiking neuron model of decision making endowed with synaptic plasticity that follows a reward-dependent stochastic Hebbian learning rule. This general model is tested in a matching task in which rewards on two targets are scheduled randomly with different rates. Our main results are threefold. First, we show that plastic synapses provide a natural way to integrate past rewards and estimate the local (in time) "return" of a choice. Second, our model reproduces the matching behavior (i.e., the proportional allocation of choices matches the relative reinforcement obtained on those choices, which is achieved through melioration in individual trials). Our model also explains the observed "undermatching" phenomenon and points to biophysical constraints (such as finite learning rate and stochastic neuronal firing) that set the limits to matching behavior. Third, although our decision model is an attractor network exhibiting winner-take-all competition, it captures graded neural spiking activities observed in LIP, when the latter were sorted according to the choices and the difference in the returns for the two targets. These results suggest that neurons in LIP are involved in selecting the oculomotor responses, whereas rewards are integrated and stored elsewhere, possibly by plastic synapses and in the form of the return rather than income of choice options.

Key words: **matching behavior; reward-dependent stochastic Hebbian learning; lateral intraparietal cortex; melioration; decision making; dopamine**

Introduction

In natural behavior, what often matters is how we make a series of choices over time, rather than an isolated decision. For example, the success in foraging for nourishment depends on the temporal pattern of food gathering; one's diet is determined by how frequently one selects food alternatives over a period of time. For decades, psychologists have studied individuals' allocation of repeated responses to a set of choices in laboratory experiments using foraging-type tasks. In these tasks, the environment is uncertain and the same choice can lead to different outcomes (no reward, or reward of varying magnitude); thus, decision making is inherently probabilistic. These studies have led to Herrnstein's "matching law," which states that a subject allocates her or his choices in a proportion that matches the relative reinforcement obtained on these choices (Herrnstein, 1961; Williams, 1988; Herrnstein et al., 1997). The matching law has been shown to be valid in a variety of task paradigms, and across species (e.g., pigeons, rats, monkeys, humans) (de Villiers and Herrnstein, 1976; Williams, 1988; Gallistel, 1994; Anderson et al., 2002).

Matching law is about an individual's choice, hence ultimately should be explained in terms of neural processes of decision making in the brain. Recently, neurobiologists have embarked on this quest and have begun to identify single neuronal activities in the primate brain that are correlated with matching behavior. In particular, several studies have used oculomotor tasks, in which typically two visual targets for saccadic eye movements are associated with different probabilities and/or magnitudes of rewards (Platt and Glimcher, 1999; Sugrue et al., 2004a; Lau and Glimcher, 2005b). Platt and Glimcher (1999) found that spike firing of single cells in the lateral intraparietal cortex (LIP) (a cortical area critical to oculomotor behavior) varies with the relative gain that an animal expects from each response, as well as with the probability of obtaining such a reward. Therefore, it was suggested that the LIP activity is modulated by relative profitability (expected gain times reward frequency). Sugrue et al. (2004a) used concurrent variable-interval schedules similar to the original Herrnstein experiment, and found that activities of some LIP neurons were correlated with a representation of value that the authors defined as fractional income. This study along with other studies (Platt and Glimcher, 1999; Dorris and Glimcher, 2004) indicates that LIP neurons reflect the values of possible actions, although these values are likely to be computed elsewhere in the brain [LIP neurons are selective to the spatial location of a visual target, whereas in this experiment the value (baiting probability) is associated with target color rather than its location]. Moreover, phenomenological models, in which the local (response by response) decision is based on time integration of past rewards (Sugrue et al., 2004a; Corrado et al., 2005b) or both past rewards and choices

Received Dec. 3, 2005; revised Jan. 26, 2006; accepted Feb. 17, 2006.

This work was supported by National Institutes of Health Grants MH062349 and MH073246 and by the Swartz Foundation. We are grateful to Stefano Fusi for the early work and helpful discussion on stochastic learning and to L. Sugrue for helpful discussions on the matching experiment. We also thank Z. Ayubi, R. Gallistel, B. Lau, D. Luxat, and P. Miller for helpful comments on this manuscript.

Correspondence should be addressed to Xiao-Jing Wang at the above address. E-mail: xjwang@brandeis.edu.

DOI:10.1523/JNEUROSCI.5159-05.2006

 0270-6474/06/263731-14$15.00/0

(Lau and Glimcher, 2005b) have been shown to reproduce monkeys' matching behavior. These and other models (Williams, 1988; Gallistel et al., 2001), however, do not address the question of what cellular and circuit mechanisms, at the biophysical level, underlie the matching behavior, which is the subject of the present work.

Our starting point is a biophysically based spiking neuron model of decision making (Wang, 2002) that has been shown to capture psychological behavior and corresponding LIP neural activities in perceptual (visual motion) discrimination tasks (Shadlen and Newsome, 1996, 2001; Roitman and Shadlen, 2002). In that model, two groups of neurons (tuned to different targets) integrate inputs over time, and the choice is selected according to which of the two neural groups wins the competition.

In the present study, we incorporated reward-dependent synaptic plasticity into our neuronal decision-making model. Specifically, we used binary synapses that undergo a stochastic Hebbian learning rule (Amit and Fusi, 1994; Fusi, 2002), with the additional condition that coactivation of presynaptic and postsynaptic neurons leads to potentiation only if the choice is rewarded, and depression otherwise. This was inspired by the suggestion that the presence or absence of dopamine signal modulates the synaptic plasticity at corticostriatal and prefrontal synapses (Reynolds et al., 2001; Reynolds and Wickens, 2002; Jay, 2003; Otani et al., 2003; Huang et al., 2004). Our working hypothesis is that input synapses onto a decision circuit (like LIP) are updated at the end of each trial according to such a reward-dependent Hebbian learning rule. As a result of synaptic modifications, the difference in the input strengths for the two competing neural groups of the decision network varies from trial to trial, which leads to adaptive dynamics of choice behavior.

Our model endowed with plastic synapses is a general one, not designed specifically for a particular behavioral task. In this paper, we report model simulations in which the two competing choices were rewarded stochastically at different rates, like in a matching task (Sugrue et al., 2004a; Lau and Glimcher, 2005b). We found that the model reproduces the neurophysiological as well as behavioral observations from the monkey experiment (Sugrue et al., 2004a). We show that plastic synapses provide a natural mechanism for computing local returns (local time average of reward per choice). Moreover, the model operates in single trials according to the so-called melioration principle (i.e., in each trial, the decision is biased toward the choice with a higher return) (Herrnstein and Vaughan, 1980; Vaughan, 1981; Herrnstein and Prelec, 1991), which ultimately gives rise to the global matching behavior. Some preliminary results have been reported in abstract form (Soltani and Wang, 2004).

Materials and Methods

Decision-making network. The decision-making network model used here is the same as the one in the study by Wang (2002); all of the model details can be found therein (see also Brunel and Wang, 2001). Briefly, the model consists of 2000 integrate-and-fire (1600 excitatory and 400 inhibitory) neurons, which are organized into three populations of excitatory neurons (two selective for competing targets, A and B, whereas the third one is nonselective) and one single population of inhibitory neurons. Recurrent synaptic currents are modeled by realistic kinetics, mediated by AMPA and NMDA receptors for excitation and by $GABA_A$ receptors for inhibition. In addition to recurrent synaptic inputs from other neurons in the network, every neuron receives independent background and afferent excitatory inputs (mediated by AMPA receptors) from 800 presynaptic neurons outside the network. The background presynaptic neurons fire constantly at 3 Hz, and these external spikes are generated with Poisson statistics.

With the presentation of visual targets, neurons in both selective populations receive a combination of two inputs (mediated by AMPA receptors). The first one, through a feedforward sensory pathway, codes the target appearance, whereas the second one codes the target color via an indirect pathway. Specifically, the first input (identical for both targets) is mediated by some afferent neurons (eight presynaptic neurons for each neuron in the selective populations), which after the appearance of the two targets, increase their firing rates from 3 to 55 Hz. Moreover, these neurons exhibit spike-frequency adaptation with a time constant of 120 ms and a steady-state firing rate of 8 Hz. In this way, neurons in the two selective populations display an initial peaked response followed by a decay to a lower steady-state response, similar to the response of LIP neurons after the onset of two targets (Sugrue et al., 2004a). The second input is mediated by some other afferent neurons (four presynaptic neurons for each neuron in the selective populations), which increase their firing rates from 1 to 10 Hz during the target presentation. Because in the experiment of Sugrue et al. (2004a), rewards were associated with the choice about color (red and green), we assume that synapses of the second pathway are endowed with reward-dependent plasticity. Because the second input is presumed to arrive through an indirect pathway passing several synapses, we added a latency of 50 ms between the onset of the first and second inputs (Schmolesky et al., 1998).

In the experiment of Sugrue et al. (2004a), the target color (red and green) encodes the rewarding value of each target, but the location of red and green targets was randomized from trial to trial. So in order for spatially selective LIP neurons to receive the correct information about the rewarding value of the leftward and rightward targets, a remapping from color to location should take place in each trial. The model presented in this paper does not explicitly address the issue of remapping from color to location (because we do not have enough experimental information yet on which to build a realistic model), but in the supplemental material (available at www.jneurosci.org), we describe a schematic circuit that is potentially able to perform such a remapping. Regardless of the details of implementation, what is essential for our model is the assumption that competing neural populations in the decision-making network are selective to target options (A or B) and receive inputs that convey information about the associated rewards via plastic synapses.

For the sake of simplicity, we did not model an additional network that reads out the decision choice. Instead, we assumed that 1.2 s after the target onset, if the difference between the average firing rates of the two (A and B) selective populations exceeded a fixed threshold of 8 Hz (for an interval which lasts $>$50 ms), then the choice of the network was the population with a higher firing rate. In the rare situation when this criterion was not met until 1.5 s after the trial onset, the threshold for making a decision was lowered to 4 Hz. After the decision is made, a decrease in the input firing rates brings the network to a regime that only one of the selective populations can stay at a high level of activity, so at the time of reward delivery the activity in only one of the selective populations is high.

Reward-dependent plasticity rule. The plastic synapses of the second input pathway are assumed to be binary (Petersen et al., 1998; O'Connor et al., 2005), with two discrete states: a potentiated "Up" state with peak conductance of $g_+ = 5.5$ nS, and a depressed "Down" state with peak conductance of $g_- = 0.5$ nS. At any moment, a fraction c of these synapses are in the Up state, whereas the remaining fraction, $1 - c$, are in the Down state. Plasticity is implemented by activity-dependent modifications of c_A and c_B for the two selective and competing neural populations.

The learning rule we use has three characteristics (Fusi et al., 2005a). First, it is Hebbian, depending on the firing rates of presynaptic (target-coding) and postsynaptic (decision) neurons. Second, an all-or-none reward signal (depending on the outcome of a target selection) can reverse the direction of plasticity (potentiation if reward is harvested, depression otherwise). Third, when the Hebbian condition is met, synaptic modification occurs probabilistically (Amit and Fusi, 1994; Fusi, 2002; Fusi et al., 2005b). In potentiation instances, each synapse in the Down state has a probability q_+ to be switched to the Up state. Similarly, in depression instances, each synapse in the Up state has a probability q_- to be switched to the Down state.

Based on these rules, the fraction of synapses in the Up state, c_i, is updated at the end of each trial as follows:

$$c_i(n+1) = c_i(n) + q_+(r;\nu_i)[1 - c_i(n)] \text{ in the case of LTP} \quad (1)$$

$$c_i(n+1) = c_i(n) - q_-(r;\nu_i)c_i(n) \text{ in the case of LTD,} \quad (2)$$

where i = A or B, and $q_+(r;\nu_i)$ and $q_-(r;\nu_i)$ are the potentiation and depression rates, respectively (termed together as learning rates). The second term in Equation 1 describes the change attributable to the transition of synapses in the Down state, because a fraction $1 - c_i$ of synapses are potentiated with probability $q_+(r;\nu_i)$. Similarly, the second term in Equation 2 describes the change attributable to the transition of synapses in the Up state, because a fraction c_i of synapses are depressed with probability $q_-(r;\nu_i)$. The learning rates depend on the firing rate ν_i, of the postsynaptic decision neurons at the end of each trial, and on the outcome of the decision r, which is either rewarded or unrewarded. The firing rate ν_i is low for the neurons selective to the unchosen target, and it is high for the neurons selective to the chosen target. For most of the results presented in this paper, unless stated otherwise, we assume that the depression and potentiation rates are constant and nonzero if ν_i is high, so synaptic plasticity only happens to the set of synapses projecting to the winning neural population. In this case, the learning rule simplifies to the following:

$$c_i(n+1) = c_i(n) + q_+[1 - c_i(n)] \text{ target } i \text{ is selected and rewarded}$$
$$c_i(n+1) = c_i(n) - q_- c_i(n) \text{ target } i \text{ is selected but not rewarded.} \quad (3)$$

Matching task simulation. An oculomotor matching task paradigm similar to that of Sugrue et al. (2004a) was simulated. In this task, a monkey was trained to choose between two visual targets with different colors (red and green). A selection of each target is rewarded independently and stochastically at a certain rate (with Poisson statistics). Reward in this task was persistent in the sense that, if a reward was assigned to a target, it stayed there until it was harvested. To discourage the monkey from switching between the two targets, a change-over-delay (COD) penalty was imposed, so if the monkey switched from one target to the other, it should choose the new target for the second time to harvest any baited reward on it. The probability of baiting rewards on the two targets (reward schedule) changed between blocks of trials without any warning to the monkey. The baiting probability ratios were randomly chosen from the ratios [1:1, 1:3, 1:6, 1:8], whereas the overall baiting probability was fixed.

In our study, we use a discrete version of the same task so in each trial if a target was not baited with a reward, the computer assigned a reward to that target with some probability, independently of the other target. The overall baiting probability is set to 0.3 rewards per trial to match the reward rate in the experiment of Sugrue et al. (2004a). A sequence of blocks with different baiting probability ratios is called a "session." In most of the simulations, the model encountered a session of the matching task with baiting probability ratios [1:1, 1:3, 3:1, 1:1, 3:1, 1:3, 1:1, 1:6, 6:1, 1:1, 6:1, 1:6, 1:1, 1:8, 8:1, 1:1, 8:1, 1:8, 1:1], which were presented in a sequence of blocks of trials. This reward schedule was usually fixed when the performance of the model was assessed with a range of parameter values. In this way, the model has been tested with the most drastic changes in the reward schedule. The average choice behavior of the model is then computed using all blocks of trials. As observed in the experiment, monkeys obey the COD constraint so most of the time they stay on the new selected target after a switch. In our modeling, we impose the COD by requiring the model to choose the same target after any switch. Similar to other trials, in the trials after switches, plastic synapses undergo changes according to the same learning rule. In one simulation (see Fig. 6), we relax the COD constraint so in that case there is no mandatory movement after a switch and the model freely chooses between the two targets in each trial.

It is important to define the following terms that we use throughout the paper. If from the total number of N trials, N_A of them were choices for target A and N_B were choices for target B, then the probability of choosing A, P_A, is equal to N_A/N. At the same time, if the total M_A and M_B rewards have been harvested on target A and B, then the incomes from target A and B, I_A and I_B, are equal to M_A/N and M_B/N, respectively. Furthermore, returns from target A and B, R_A and R_B, are equal to M_A/N_A and M_B/N_B, respectively.

Results

Behavior of the decision-making network and emergence of graded activity

The behavior of our model results from an interplay between the decision process and synaptic plasticity. In any trial, given the synaptic strengths c_A and c_B, the network integrates inputs and makes a choice. At the end of each trial, depending on the choice and whether it is rewarded, c_A or c_B is updated, which in turn influences the decision process in the subsequent trial. We first quantify the decision process of the network as a function of fixed c_A and c_B values. Then we will consider trial-to-trial modifications of c_A and c_B and the resulting dynamic decision making over time.

In our simulation of the experiment of Sugrue et al. (2004a), it is reasonable to assume that the two visual targets lead to identical firing rates of presynaptic sensory neurons that project to our decision network. Therefore, the only difference in the inputs to the two competing neural populations is attributable to the efficacies of the plastic synapses. The average synaptic conductance of input plastic synapses is a function of multiple factors, including the number of plastic synapses, the presynaptic firing rate, and the peak conductance of the potentiated and depressed states, and can be written as follows:

$$G = N_p f_{st}(c g_+ + (1 - c) g_-)\tau_{syn}, \quad (4)$$

where N_p is the number of plastic synapses onto each neuron, f_{st} is the firing rate of the presynaptic neurons, g_+ and g_- are the peak conductance of the synapses in the Up and Down states, respectively, and τ_{syn} is the decay time of AMPA currents. Thus, the difference in the average synaptic conductances of neurons in the two selective populations, can be quantified as a function of the synaptic strengths (fraction of synapses in the potentiated state) c_A and c_B.

$$G_A - G_B = (c_A - c_B) N_p f_{st}(g_+ - g_-)\tau_{syn}. \quad (5)$$

As we show later, the choice behavior of the network is a function of the difference in synaptic strengths $c_A - c_B$ (or equivalently $G_A - G_B$), so the multiplicative factor $N_p f_{st}(g_+ - g_-)$ can change the sensitivity (which we will call σ) of the choice behavior to the difference in synaptic strengths.

The behavior of the network in 10 sample trials, with fixed $c_A = 0.33$ and $c_B = 0.27$, is illustrated in Figure 1. As is evident, at the onset of the targets, the firing rates of the two neural populations initially increase together for a few hundred milliseconds, and then start to diverge so that firing rate of one population (e.g., A) keeps increasing while the firing rate of the other population (e.g., B) decreases gradually.

This "winner-take-all" competition process is attributable to effective mutual inhibition (through the shared interneuron pool) between neural populations A and B. Consequently, at the end of a trial, a categorical choice can be read out according to which of the two neural populations has a higher firing rate. In the model, the high level of activity in the winning population can be self-sustained, even after the removal of the stimulus, because of recurrent reverberation. Hence, in principle, the choice can still be read out after a memory delay period (Wang, 2002).

If the difference between c_A and c_B is not too large, the decision of the network is probabilistic, because neural spike discharges

are intrinsically stochastic (note the trial-by-trial variability of population firing rates in Fig. 1). For example, with $c_A = 0.33$ and $c_B = 0.27$, the network chooses A in 78% of trials and B in 22% of trials (sample trials are shown in Fig. 1). A comparison between the left and right panels in Figure 1 reveals that a few hundred milliseconds after the onset of the targets, the firing rate of the winning population is somewhat lower when its synaptic strength is smaller. Moreover, the time it takes for the two neural populations to diverge (hence the "decision time") is longer and more variable.

The trial-averaged population activities are shown in Figure 2*A* for which the average activity in the two selective populations are sorted according to the choice of the network in each trial. It is apparent that a few hundred milliseconds after the onset of the targets, the activity of neurons is significantly higher when the chosen target is the preferred target (red) than when it is the nonpreferred target (blue). Furthermore, there is a graded change of activity levels between the cases in which the synaptic strength for the winning population is larger (thick curves) or smaller (thin curves) than that for the losing population.

These characteristics are robust in our model, as long as the overall synaptic strength ($c_A + c_B$) is reasonably low, which favors winner-take-all competition. As shown in Figure 2, *B* and *C*, with the same value for the difference $c_A - c_B = 0.06$ but a higher value for $c_A + c_B$ (1 and 1.4 instead of 0.6), the winning and losing neural populations still exhibit some differences that are no longer of a categorical character. This is because the external drive is now large for both neural populations and becomes predominant over the winner-take-all recurrent network dynamics, so the activity of the losing population is larger and closer to that of the winning population. In this case, the choice is determined in simulations by the neural population with a higher firing rate at the end of each trial (see Materials and Methods).

Interestingly, for all three cases in Figure 2 with $c_A - c_B = 0.06$, the choice probability turns out to be approximately the same ($P_A = 0.78$, 0.81, and 0.77, for $c_A + c_B = 0.6$, 1, and 1.4, respectively). This holds true for other $c_A - c_B$ values. As shown in Figure 3, the choice probability as a function of $c_A - c_B$ is not sensitive to the overall synaptic strength $c_A + c_B$. We fitted the probability of choosing target A by a sigmoid function of $c_A - c_B$.

$$P_A = \frac{1}{1 + \exp\left(-\frac{c_A - c_B}{\sigma}\right)} \quad (6)$$

Note that the σ value, which determines the randomness in the choice behavior of the network, depends on factors that can change the difference in the overall synaptic currents through the plastic synapses (Eq. 5).

As a result, the value of σ can be adjusted by the number of the plastic synapses, by the firing rates of presynaptic neurons projecting to the decision network, and by the peak conductance of plastic synapses (Eq. 5). For model parameters used in our simulations, we obtained σ equal to 4.84% (if $c_A - c_B$ is expressed as percentage).

It is an important feature of our model that the choice behavior is only a function of the difference ($c_A - c_B$), which we will refer to as "differential input." In Figure 4, the average population activities are sorted according to the choice of the network and the differential input value in each trial when a range of differential input values are used in different trials. The overall synaptic strength ($c_A + c_B$) in these simulations varies from 0.4 to 1.6. Similar to what is shown in Figure 2, a graded activity emerges, which is a direct result of competition in the decision-making network and difference in synaptic strengths of inputs to the two populations.

The activity of neurons, in trials in which they win the competition (the chosen target is their preferred one), is higher when the difference in the synaptic strength in their favor is greater (compare the red curves). In contrast, in trials in which they lose the competition (the chosen target is the nonpreferred one), the activity of neurons is higher when the difference in the synaptic strength in their disfavor is smaller (compare the blue curves).

Therefore, by sorting neural firing rates across trials according to the differential input graded activities emerge in our model. This plot (Fig. 4) is similar to Figure 5 in the study by Sugrue et al.

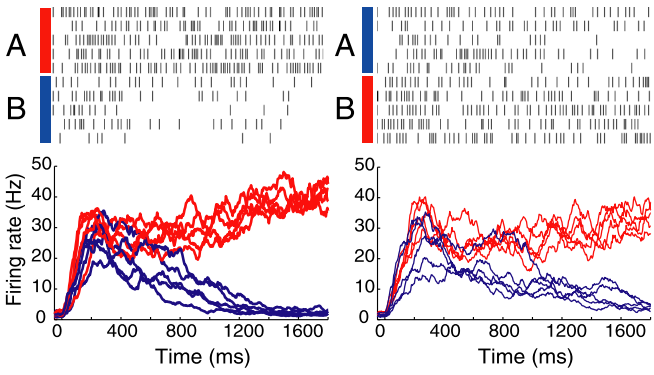

Figure 1. Neuronal activity of two selective populations of decision-making network model, in sample trials. The population firing rate of neurons is shown separately for trials in which the choice of the network is the preferred (red) or nonpreferred (blue) target of the neurons. Raster plots show spike trains for two selected neurons in populations A and B. The left panels show activity in trials in which target A is the choice of the network, and the right panels show activity in trials in which target B is the choice of the network. Activity is aligned at the onset of the visual targets. A few hundred milliseconds after the input onset, the average firing rates in the two populations start to diverge. Spiking activity is higher when the chosen target is preferred for the neuron (compare red with blue traces) and when its input is larger (compare red traces in the left and right panels). Moreover, firing activity is higher when the chosen target is nonpreferred for the neuron that receives a larger input (compare blue traces in the left and right panels). In these simulations, the synaptic strengths are $c_A = 0.33$ and $c_B = 0.27$.

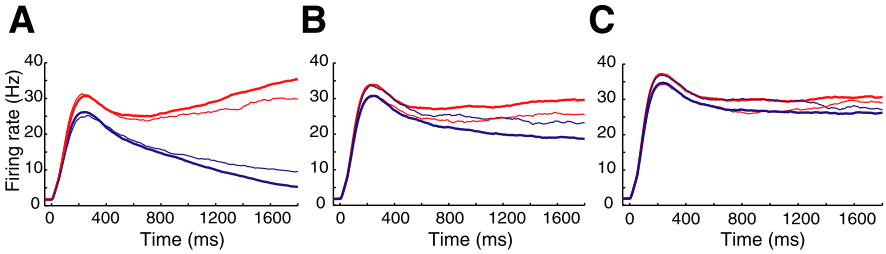

Figure 2. Average activity of the two selective populations sorted according to the choice of the network in each trial. Average activity is shown for three different sets of synaptic strengths: $c_A = 0.33$, $c_B = 0.27$ (***A***); $c_A = 0.53$, $c_B = 0.47$ (***B***); and $c_A = 0.73$, $c_B = 0.67$ (***C***). Note that, in these three cases, the differential input is the same ($c_A - c_B = 0.06$), but the overall inputs are different ($c_A + c_B = 0.6$, 1.0, and 1.4, respectively). The average activity of neurons in the two selective populations are sorted based on the choice of the network in each trial and then averaged (over 400 trials for each set of synaptic strengths). Red (blue), The choice of the network is the preferred (nonpreferred) target for the neural population. Thick (thin), The neural population selective for the chosen target receives a larger (smaller) input than its competitor. Regardless of whether the chosen target is preferred (red curves) or nonpreferred (blue curves), the average population activity is higher when the neurons receive a stronger input (compare thick and thin curves).

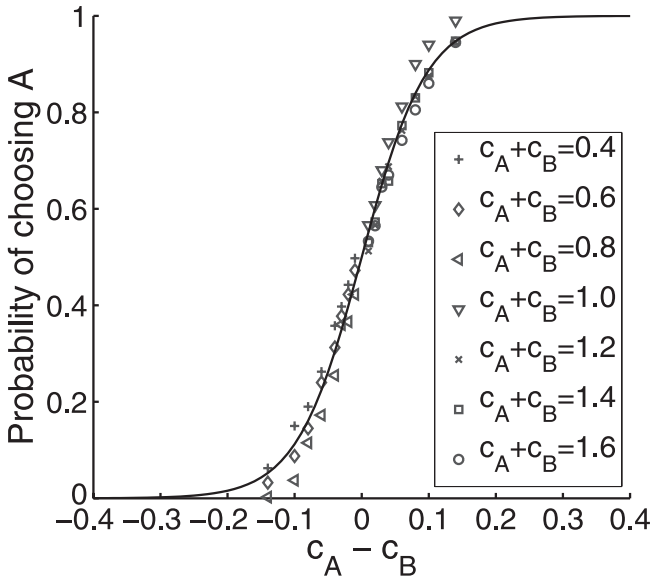

Figure 3. Choice behavior of the decision-making network for different sets of synaptic strengths. The probability of choosing target A is only a function of the difference between the two synaptic strengths, and it can be fitted by a sigmoid function. The solid curve shows the fitting by a sigmoid function ($\sigma = 4.84\%$). The choice probability for each set of synaptic strengths is obtained from 400 simulated trials. For each set of synaptic strengths with equal overall synaptic strengths, differential inputs are set to 0.01, 0.02, 0.03, 0.04, 0.06, 0.08, 0.1, and 0.14.

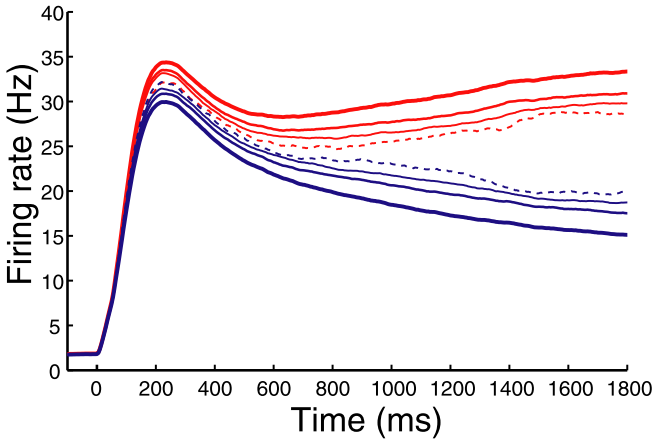

Figure 4. Graded activity of neurons in the two selective populations. The activity of decision neurons shows a graded pattern if single-trial firing rates are sorted and averaged according to the choice of the network and the difference between synaptic strengths. Activity is aligned by the onset of two targets, and it is shown separately for the choice that is the preferred (red) or nonpreferred (blue) target of the neurons. In addition, trials are subdivided into four groups according to the difference between the strength of synapses onto the two competing neural populations [$c_A - c_B = -0.05$ to -0.14 (dashed), 0 to -0.05 (thin), 0–0.05 (normal), 0.05–0.14 (thick)]. The overall synaptic strength, $c_A + c_B$, varies from 0.40 to 1.6. For these simulations, 56 different sets of synaptic strengths are used (the same values used for Fig. 3), and for each set of synaptic strengths the average activity is obtained from 400 simulated trials.

(2004a), in which LIP neural activities are sorted according to the fractional income of the chosen target, which according to the model used in that study is identical to the choice probability. Therefore, our model semiquantitatively reproduces the LIP neural spike activities reported in the matching task (Sugrue et al., 2004a). Because the choice probability P_A (the probability of choosing target A) is a sigmoid function of differential input (Eq. 6), the same kind of graded activities are expected when the neural activities are sorted according to P_A (results are not shown). We shall come back to possible implications of this result to the interpretation of the observed LIP neural activities in Discussion.

The disparate timescales of the neural firing dynamics (milliseconds) and synaptic plasticity (many trials) made it difficult to simulate the full large-scale network model of spiking neurons (which requires a time resolution of 0.1 ms) sequentially across thousands of trials. To avoid this computational hindrance, from now on we use the function $P_A(c_A - c_B)$ (Eq. 6) instead of the full spiking neural network for decision computations. Namely, in each trial, knowing c_A and c_B, we use $P_A(c_A - c_B)$ to flip a biased coin; the outcome of the coin toss determines the choice of the network in that trial (A or B). At the end of each trial, depending on the choice of the model and the presence or absence of reward, plastic synapses undergo stochastic strengthening or weakening according to Equation 3.

Learning rule and the steady state of plastic synapses

In the last section, we quantified the choice behavior of the decision-making network as a function of the differential input, $c_A - c_B$, to the two competing neural populations. Now, we consider the learning process in which c_A and c_B undergo changes depending on the decision of the network and the outcome of that decision (rewarded or unrewarded) at each trial. We first asked the question: for a given and fixed P_A (and $P_B = 1 - P_A$), how do c_A and c_B behave according to the learning rule? In particular, what are the steady-state values of c_A and c_B? In this study, we have used a simple learning rule that assumes that, at the end of each trial, only synapses projecting to the chosen population undergo stochastic strengthening (if the choice is rewarded) or weakening (otherwise).

In general, the modification rates of potentiation (q_+) and depression (q_-) can be different. In the special case in which these two learning rates are equal, the steady state of synaptic strengths of the two sets of plastic synapses are approximately equal to the returns from the two choices [this approximation holds while the learning is slow (i.e., when q_+ and q_- are small)]. This can be shown by a simple calculation (Brunel et al., 1998). The probability of obtaining a reward on target $i = A$ or B, is equal to the number of rewards on that target divided by the total number of trials, which by definition is equal to the income from target i (I_i). If the probability of choosing target i is P_i then the average change in c_i in each trial is given by the following:

$$\Delta c_i = q_+(1 - c_i)I_i - q_- c_i(P_i - I_i).$$

The first term is the change attributable to potentiation in a rewarded trial (which occurs with the probability of I_i) and the second term is the change attributable to depression in a trial in which target i is chosen, but it is not accompanied by reward (which occurs with the probability of $P_i - I_i$). In the steady state, Δc_i should be zero which results in the following:

$$c_i^{ss} = \frac{q_+ I_i}{(q_+ - q_-)I_i + q_- P_i} = \frac{q_+ R_i}{(q_+ - q_-)R_i + q_-}, \qquad (7)$$

where $R_i = I_i/P_i$ is the return from choice i (i.e., the total number of reward obtained on choice i divided by the total number of choices for i).

It is thus clear that the steady state of c_i is a function of the return, R_i, from the choice i. In the special case in which $q_+ = q_-$, the steady state of c_i is equal to R_i. Even when q_+ and q_- are different, c_i is approximately a linear function of the return, $c_i^{ss} \simeq (q_+/q_-)R_i$, as long as $|q_+ - q_-|R_i$ is much smaller than q_-. The latter inequality generally holds when the return is signifi-

cantly smaller than 1 (which is the case in the simulated experiment) and q_+ is not much larger than q_-.

Matching through probabilistic melioration

We have seen that, in our model, the return from each choice is represented in the synaptic strength of plastic synapses. The difference between synaptic strengths, $c_A - c_B$, determines the choice probability, which in turn modifies the returns. This interplay between the synaptic strengths (or equivalently returns) and the choice probability underlies dynamic decision of our model. Here, we show how this interaction can result in matching behavior. We shall first analyze an ideal situation to gain an intuitive understanding, and then consider more realistic simulations in the next subsection.

For simplicity, we focus on a discrete version of the concurrent variable-interval schedule (VI–VI), without imposing a change-over-delay penalty (see Materials and Methods). In the following, we shall begin by assuming that the model selects between the two targets with a given (current) choice probability. Based on the current value of choice probability, we then compute the returns (hence the synaptic strengths) and use Equation 6 to calculate the "predicted" choice probability from the model. We assess whether matching is achieved in the steady state when the predicted and current choice probabilities are equal (self-consistent). Specifically, for a given (current) choice probability, say for target A, P_A, the return from each target (R_A and R_B) can be computed easily (Heyman and Luce, 1979; Houston and Sumida, 1987). In Figure 5*A*, these returns are plotted (red and green curves) as a function of P_A, for given baiting probabilities. On the same plot, the total income from the two targets, $P_A R_A + P_B R_B = I_A + I_B = I_{tot}$, is plotted (blue curve).

When the returns from the two targets are equal, the total income is maximal (at $P_A = 0.782$). Therefore, in this task, matching corresponds to optimal behavior (Staddon and Motheral, 1978).

Now, in our model, the choice probability is a function of the difference between synaptic strengths, $c_A - c_B$ (Eq. 6). Assuming that the synaptic strengths are equal to the returns (red and green curves), the choice probability predicted by the model can be calculated according to Equation 6. This is plotted in Figure 5, *B* and *C* (black curve), for two different values of σ. At the intersection of the red and green curves ($P_A = 0.782$) where the returns from the two targets are equal, and hence $c_A = c_B$, the predicted choice probability by the model is equal to 0.5. If the return from target A is larger than the return from target B (for $P_A < 0.782$), the predicted choice probability is biased toward target A (>0.5) and when the return from target B is larger than the return from target A (for $P_A > 0.782$), the predicted choice probability is biased toward target B. In this sense, our model acts according to the melioration principle (Herrnstein and Vaughan, 1980; Vaughan, 1981; Herrnstein and Prelec, 1991), which states that the choice behavior should be biased toward the option with the higher return. However, decision is not deterministic in our model; the target with the higher return is chosen simply with some probability larger than 0.5.

Because of the probabilistic nature of our decision-making model, there is always a limit for approaching perfect matching. This concept is illustrated in Figure 5*B*. If the predicted choice probability computed by the model is greater than the current value of choice probability, P_A, then the model has a tendency to increase the probability of choosing target A. If the predicted choice probability is smaller than the current value of choice probability, then the model has a tendency to decrease the probability of choosing target A. The final state of the model (steady state) is the point at which the current and predicted choice probabilities are equal, that is, the intersection of the black curve and the diagonal line in Figure 5*B* ($P_A \approx 0.7$). This mechanism makes the model reach a choice probability that is generally smaller but

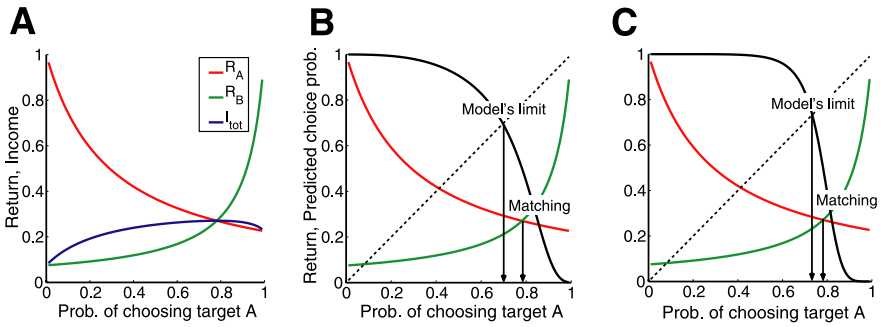

Figure 5. Mechanism of melioration and the limit for achieving perfect matching. ***A***, For a given choice probability, the return from each target (red for A and green for B) is computed and is plotted as a function of the probability of choosing target A (in all panels). The baiting probability on target A is three times the baiting probability on target B, and the overall baiting probability is equal to 0.3. Matching happens at a choice probability ($P_A = 0.782$) for which the returns from the two targets are equal. At this choice probability, the total income is optimal (blue curve) showing that, in this task, matching is optimal. ***B***, The choice probability for target A, predicted by the model (using Eq. 6) is shown in black for $\sigma = 10\%$. The steady state of the model is the point at which the predicted and current choice probabilities are equal, which is given by the intersection of the black curve with the diagonal line. The location of the steady state falls short of the prediction of matching, a phenomenon called undermatching. ***C***, For a smaller value of $\sigma = 5\%$, the steady state is closer to the prediction of matching (compare ***B***, ***C***).

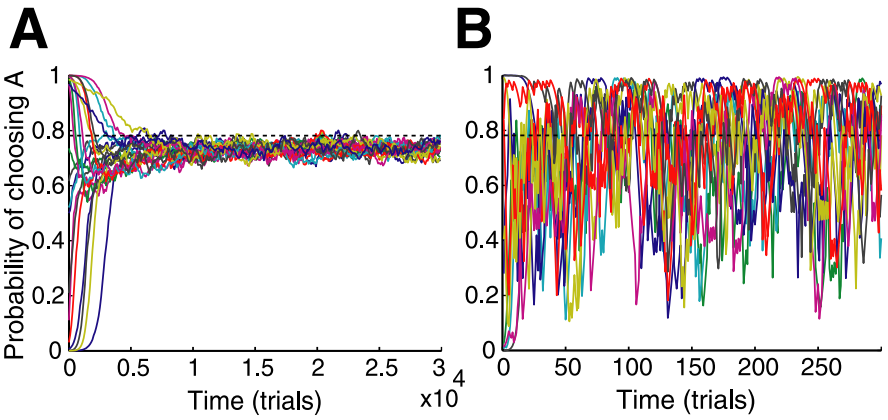

Figure 6. Time course of the choice behavior of the model in a matching task without COD penalty. The probability of choosing target A as function of time is plotted for different random initial values of c_A and c_B. The baiting probability ratio is similar to Figure 5 (3:1 in favor of target A). ***A***, The model reaches the steady-state choice behavior over a long time when the learning rates are very small ($q_+ = q_- = 0.0006$). ***B***, For more realistic values of learning rates ($q_+ = q_- = 0.06$), the model reaches the steady state quickly, but the fluctuations in the choice behavior are large. The dashed line shows the prediction by perfect matching ($P_A = 0.782$), and for both simulations σ is set to 5%.

close to that according to the matching law (0.782), a phenomenon called "undermatching." The extent of undermatching depends on the value of σ, so that, for a smaller value of σ, the steady state of the model is closer to the prediction of matching ($P_A \approx 0.73$ in Fig. 5*C* instead of $P_A \approx 0.7$ in Fig. 5*B*).

We found that this steady state of the model is stable, that is, the final state of the model is the same independent of the initial condition. Examples of the model choice behavior as a function of time are shown in Figure 6*A*, for which the learning rates are set to very small values. If the learning rates are large, the choice probability converges to a steady state fast, but fluctuations around the steady state are large (Fig. 6*B*).

So far, we have discussed the case in which the potentiation and depression rates are the same, so the synaptic strengths are equal to the returns. What happens if the rates of potentiation and depression are not equal? As we mentioned, if the overall baiting probability is small and q_+ is not much larger than q_-, c_A and c_B are still linear functions of returns from the two choices. For given R_i values, if $q_+ > q_-$, the differential input ($c_A - c_B$) is larger than the difference between returns (because the slope of c_i values as a function of return is >1). Because the choice probability P_A is a function of the ratio of the differential input to σ (Eq. 6), a larger differential input is equivalent to an effectively smaller value of σ, which results in better matching. If $q_+ < q_-$, the differential input is smaller than the difference in returns, which is equivalent to a larger value of σ; thus, the model shows more undermatching (data not shown).

Choice behavior of the model in a dynamic environment

In the previous subsection, we established that in a stationary environment for which the baiting probabilities stay constant, our model is able to reach a choice behavior close to matching. Now, we investigate whether this holds true in a dynamic environment, especially when the baiting probability ratio changes frequently between blocks of trials.

We simulate a matching task experiment in which the baiting probability ratio changes between blocks of 200 trials, similar to the task used by Sugrue et al. (2004a). An example of the choice behavior of the model in one simulated session of the experiment is shown in Figure 7*A*.

In each block of trials, the choice ratio, which is the slope of the cumulative choice plot, approximately matches the baiting probability ratio (Fig. 7*A*, black straight lines). Moreover, the instantaneous choice fraction closely follows the instantaneous reward fraction (Fig. 7*C*), an indication that matching is achieved dynamically in our model. The systematic trial-to-trial change in the choice behavior is determined by the ongoing changes in the synaptic strengths. To show these changes, synaptic strengths and choice probability are plotted during the same simulation (Fig. 7*D*). In each trial, only the synaptic strength that corresponds to the selected target undergoes small modification, and this modification is enough to alter the choice probability in each trial. In contrast, in each block of trials, the synaptic strengths fluctuate around different average values depending on the baiting probabilities in that block. In the example shown, these average values are the returns from each target. To demonstrate this point more clearly, we plot the average synaptic strengths in each block of trials versus the return from the corresponding choice in the same block (Fig. 8*A*).

Similar to the analytical prediction, if the two learning rates are equal, the averaged value of synaptic strengths in each block, $\langle c_A \rangle$ and $\langle c_B \rangle$, are close to the returns from choice A and B, respectively. In the cases in which the learning rates are not equal (Fig. 8*B*), average synaptic strength is still approximately a linear function of the return.

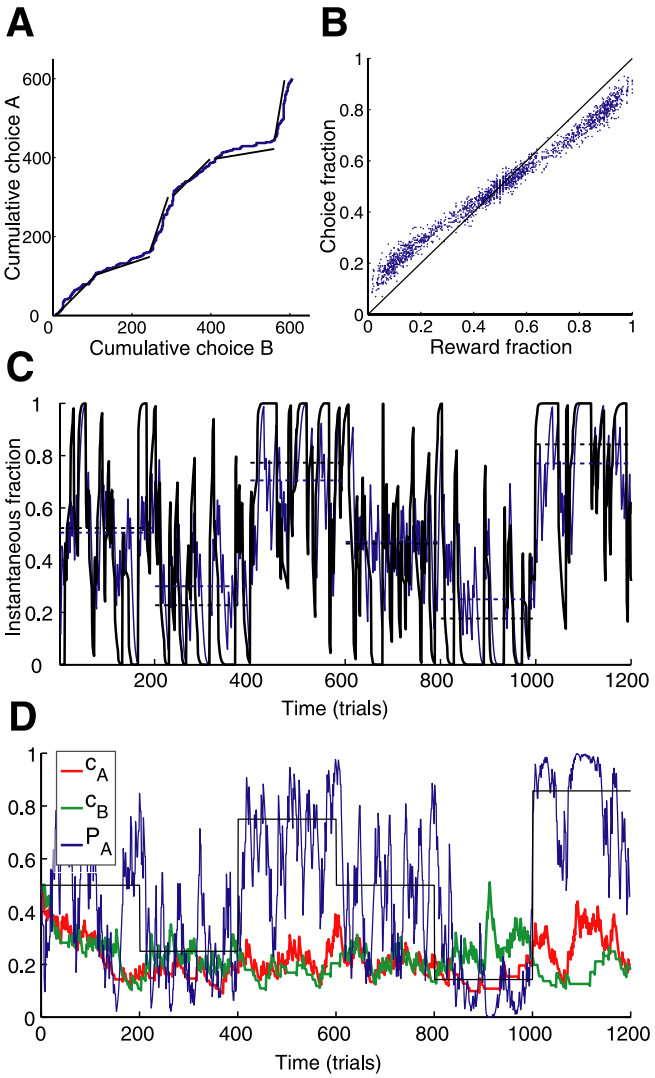

Figure 7. Model shows matching in a dynamic environment. ***A***, For one session of the simulated matching experiment, the cumulative choice on target A is plotted versus the cumulative choice on target B. The black straight lines show the baiting probability ratio in each block. The slope of the cumulative plot is equal to the choice ratio and is approximately equal to the baiting probability ratio. In this session the following baiting probability ratios are used in sequence [1:1, 1:3, 3:1, 1:1, 1:6, 6:1]. ***B***, Each point shows the blockwise choice fraction as a function of the blockwise reward fraction for a block of trials on which the baiting probabilities are held constant. The baiting probability ratios are selected from all possible ratios (see Materials and Methods). Most of the points fall close to the diagonal line (perfect matching), but the choice fraction is slightly lower than the reward fraction when the latter is larger than 1/2 (a phenomenon called undermatching). ***C***, The instantaneous choice (blue) and reward (black) fractions as a function of time computed for the same session shown in ***A***. The dashed lines show average choice and reward fractions for each block (in blue and black, respectively). To compute the instantaneous fractions, the choice and reward fraction are smoothed with a causal half-Gaussian filter with SD of six trials. The model is able to follow changes in the reward schedule. ***D***, The synaptic strengths, c_A (red) and c_B (green), and the choice probability (blue) as a function of time for the same data shown in ***A*** and ***C***. The thin black line indicates the baiting probability ratio in each block. In each block, the synaptic strengths fluctuate around the value of returns from the two choices. The model parameters for these simulations are $q_+ = 0.06$, $q_- = 0.06$, and $\sigma = 5\%$; and the length of each block is set to 200 trials.

To show the global choice behavior of the model, in Figure 7*B*, we plot the blockwise choice fraction (proportion of choice on target A in a block) versus the blockwise reward fraction (proportion of reward obtained from target A). The model shows good

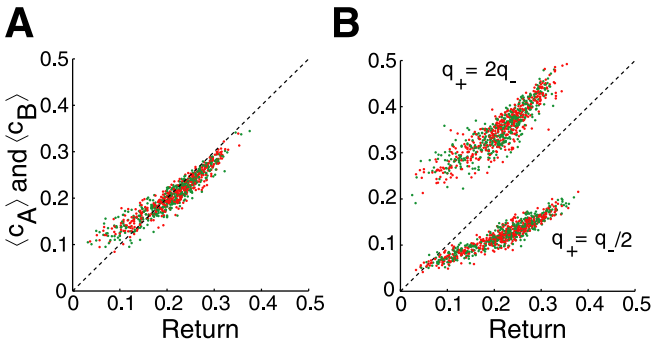

Figure 8. Plastic synapses approximately compute the return from each choice (or a linear function of it). Block-averaged synaptic strengths are plotted versus the obtained returns in the same block. Two colors represent two different choices (red for A; green for B). ***A***, The average synaptic strength is equal to the return from each choice, when the two learning rates are equal ($q_+ = q_- = 0.06$). ***B***, If the potentiation rate is greater than the depression rate ($q_+ = 0.06$; $q_- = 0.03$), the average synaptic strength is larger than the return (top points), whereas, if the depression rate is greater than the potentiation rate ($q_+ = 0.03$; $q_- = 0.06$), then the synaptic strength is smaller than the return (bottom points). These data points are obtained from 15 simulated sessions of the matching task in which all possible baiting probability ratios are used (see Materials and Methods), and the length of each block is set to 200 trials. For all simulations, σ is set to 5%.

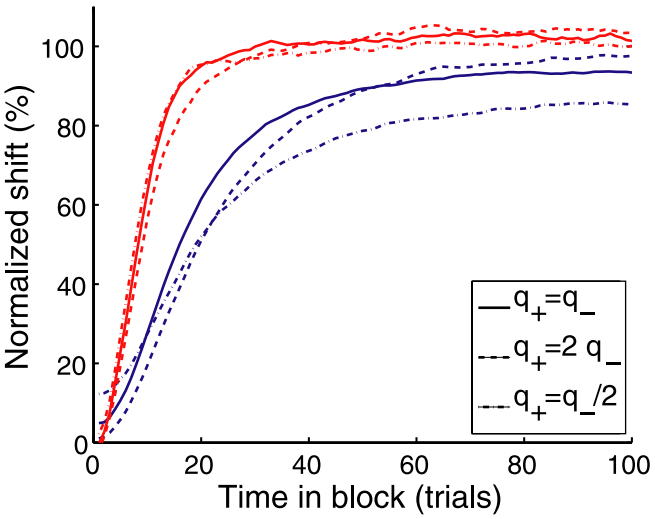

Figure 9. Adaptation to change in the reward schedule. The average time course of adjustment for the choice (blue) and reward (red) fractions in a new block of trials are plotted for three different values of learning rates: $q_+ = q_- = 0.06$ (solid curves); $q_+ = 0.06$, $q_- = 0.03$ (dashed curves); and $q_+ = 0.03$, $q_- = 0.06$ (dot-dashed curves). The choice and reward fractions are normalized, so 0% shift indicates the same fraction as the fractional baiting probability in the previous block, and 100% shift indicates the same fraction as the fractional baiting probability in the current block. The instantaneous choice and reward fractions are computed using a causal half-Gaussian filter (SD = 6 trials). The shift in the reward fractions happens in the span of 10–15 trials, and it is approximately independent of the learning rates. The shift in the choice fraction reaches an asymptotic value after 30–40 trials, which is dependent on the learning rates. If $q_+ > q_-$, the shift in choice behavior is slower right after the block transition, but its reaches a higher value later on. These results are obtained from 500 simulated sessions of matching task with all possible baiting probability ratios (see Materials and Methods). The length of each block is set to 200 trials, and σ is set to 5% in all simulations.

matching, that is, the choice fraction in each block of trials is approximately equal to the reward fraction in that block. Moreover, in blocks in which target A is richer (reward fraction >0.5) the choice fraction is usually smaller than the reward fraction. This general tendency of the model, called undermatching, has also been observed in matching task experiments in monkeys (Anderson et al., 2002; Sugrue et al., 2004a; Lau and Glimcher, 2005b).

In the previous section, we showed that, in a stationary environment, the probabilistic nature of decision making in our model results in undermatching. In a dynamic environment, undermatching becomes even more prominent, because after a change in the reward schedule, it takes a few trials for the choice behavior to be shifted according to the new reward schedule. To illustrate how fast the model is able to shift its choice behavior between blocks of trials, following the same method (with a slight modification) used by Corrado et al. (2005), we plot the normalized shift (which is the shift per trial normalized by the programmed reward fraction shift) in choice and reward fractions after each block transition (Fig. 9).

Similar to the monkey experiment, it takes ~30–40 trials for the model to completely shift its choice behavior after a block transition (Corrado et al., 2005). Furthermore, the relative values of the learning rates affect how quickly shifts can take place. For example, if $q_+ > q_-$, the shift is slower at the beginning of a block transition but reaches a higher asymptotic value later on. This happens because in this case the difference between the average value of synaptic strengths is large, and with a slow depression rate it takes more time to reverse this difference in the new block (in which the more rewarding choice is different from the last block).

Stay length and switch probability

A few behavioral studies of matching tasks in pigeons and rats have shown that switching between choices is approximately a stochastic process that only depends on the reinforcement rate on those choices (Heyman and Luce, 1979; Gallistel et al., 2001). This means that the probability that the animal switches from one choice to the other is almost independent of the time that it has spent on the current choice. Based on these results, it has been claimed that matching cannot be generated by a mechanism that involves feedback (Heyman, 1979; Gallistel et al., 2001).

We assessed the statistics of choice behavior separately for each block of trials with different baiting probability ratios. Note that for all analyses presented in this section, the mandatory movements after switches are removed because in these trials the choice behavior follows a deterministic rule. In Figure 10*A*, the distribution of stay lengths (number of consecutive choices on one target) is plotted for the two targets in different blocks of the experiment. In addition, each stay length histogram is fitted with an exponential distribution (black curves).

Consistent with the experimental observation (Corrado et al., 2005), the distribution of stay lengths on each target is approximately exponential, although a small deviation can be seen clearly. Moreover, for the target with a larger baiting probability the stay length distribution has a larger mean.

If switching between the two targets is a completely stochastic process with a rate determined by the baiting probability, one expects that probability of staying longer than a given stay length (survival probability) is an exponentially decreasing function of the stay length. As shown in Figure 10*B*, the probability of staying longer than a given stay length is approximately a linear function of the stay length in a semilog plot. The negative of the slope in this plot is approximately equal to the probability of switching to the other target (Gallistel et al., 2001). We also have computed the probability of switching as a function of the stay length (Fig. 10*C*). If the survival probability is perfectly monoexponential, then the switching probability should be independent of the stay length. We found that this is only approximately true, for long stays.

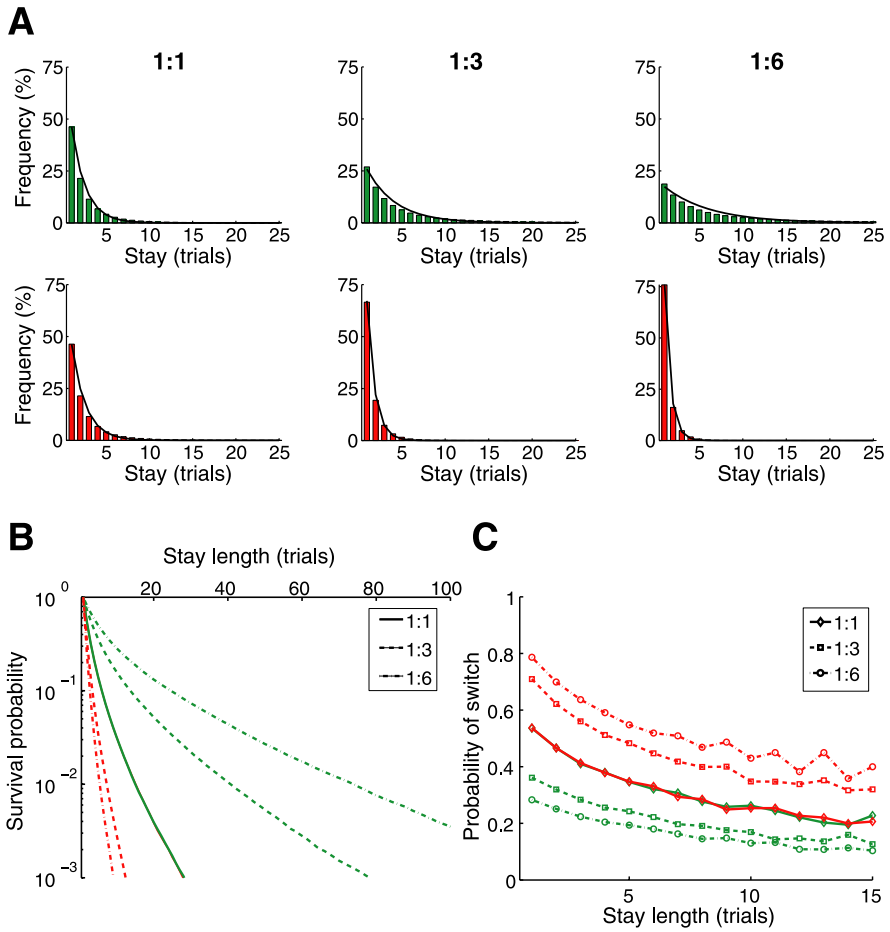

Figure 10. Statistics of the stay length and switch probability. ***A***, The distribution of stay lengths in each block (with fixed baiting probabilities) is approximately an exponential and depends on the baiting probability. The baiting probability ratio in each block is reported on each plot (in the favor of target B). For the 1:1, 1:3, and 1:6 schedule, the mean stay length for target A (red histograms) is 2.65, 1.63, 1.38, and for target B (green histograms) is 2.65, 5.71, 9.66, respectively. The black curves show the fitting with an exponential distribution. ***B***, Log cumulative probability of staying longer than a given stay length (survival probability) is plotted separately for different targets in blocks of trials with different baiting probability ratios. Red and green curves, The survival probabilities for the targets A and B, respectively. The baiting probability ratio for each block is shown in the inset. The negative of the slope in this plot is equal to the probability of switching to the other target. ***C***, The probability of switching as a function of the stay length. These statistics are obtained from 5000 simulated sessions in which the baiting probability ratio is [1:1, 1:3, 1:6] and the overall baiting probability is fixed to 0.3 (block length is equal to 2000 trials). The model parameters are $q_+ = q_- = 0.06$ and $\sigma = 5\%$.

Indeed, the probability of switching is a decreasing function of the stay length, and it reaches a steady state after ~10–15 trials. Note that the exact value of switch probability and its steady state depends on the model parameters, but its qualitative form is the same. Furthermore, the probability of switching is a function of baiting probability, so it is larger for the target with a smaller baiting probability (Fig. 10*C*).

In our model, staying on a choice or switching to another choice is a stochastic process with a probability that is determined by the state of plastic synapses in each trial (except for the mandatory movement after a switch). Note that the probability that the model stays on a target for a large number of trials is very small (especially when the baiting probability is low), and because of the stochastic nature of our model a long stay requires harvesting of a few rewards on that target. As a result, the probability of switching after a few stays decreases as a function of the stay length. Our results demonstrate that in a model based on feedback, like our model, the probability of switching is determined by the baiting probability. In addition, the slow change of the probability of switching as a function of the stay length may not be incompatible with the experimental observation of approximately constant switch probability.

Robustness of the model

Although in the previous simulations a specific set of parameters is used, matching behavior can be achieved over a wide range of parameters. The behavior of the model is quantified, using a sequence of blocks of trials with different baiting probability ratios, when the learning rates, q_+ and q_-, are varied in a broad range (with fixed σ for the noise level). We quantify the performance of the model by the ratio of the average reward rate to the overall baiting probability, as in the monkey experiment. The performance of the model is assessed with a sequence of blocks with different baiting probability ratios (see Materials and Methods) and is plotted in Figure 11*A*. For most of the potentiation and depression rates, the performance of the model is high compared with the monkeys' performance [~72% in the study by Sugrue et al. (2004a)] and does not vary significantly over the range of the parameters.

The performance is relatively low in two cases. In the first case, the potentiation rate, q_+, is much larger than the depression rate, q_- (for moderate values of q_-). This condition results in long stays on the richer target in each block (see the switching probability in Fig. 11*C*), so the model is slow in shifting its behavior between blocks of trials and loses some of the rewards. In the second case, the depression rate is much larger than the potentiation rate so any unrewarded trial gives rise to a strong reset of the synaptic strength for the chosen target. As a result, the model alternates frequently between the two choices (Fig. 11*C*). For small values of q_-, the performance of the model is high, because in these cases both synaptic strengths saturate and reach a value close to 1, and as a result the choice behavior becomes more random.

We define the "deviation from matching" as the average absolute difference between choice and reward fractions on each block. This quantity is more strongly dependent on the learning rates, although for most of the parameter space its value is small (Fig. 11*B*). Similar to the results for the choice behavior in a stationary environment, generally better matching can be achieved with a potentiation rate larger than the depression rate.

As we mention for this task when the baiting probabilities are constant, matching is optimal. However, if the baiting probability changes between blocks of trials, this statement no longer holds. In a stationary environment, good matching can be achieved when the potentiation rate is higher than the depression rate and there are not many switches between the two choices. But with these conditions, in a dynamic environment the choice behavior cannot be shifted quickly, and as a result the performance will deteriorate. So in a dynamic environment, the model parameters that result in the best matching behavior do not correspond to

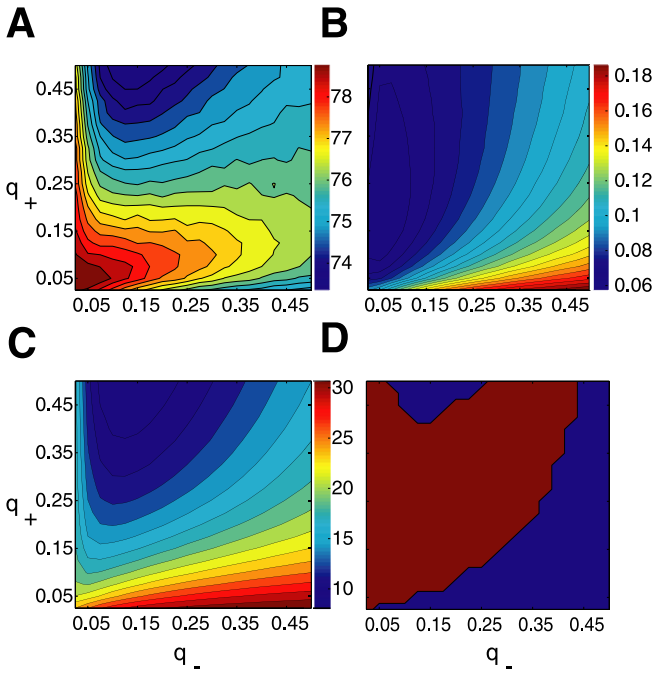

Figure 11. Model shows matching behavior over a wide range of parameters. ***A***, The performance of the model, defined as the ratio of the average reward rate to the overall baiting probability l_{tot}/λ_{tot}, is generally high and does not change significantly with learning rates (only a few percent change). ***B***, The "deviation from matching," computed as the average of absolute difference between choice and reward fractions on each block, is small over a wide range of learning rates. This indicates that, for a wide range of learning rates, a choice behavior close to matching can be achieved. ***C***, The switching probability (expressed in percentage), the total number of switches between the two choices divided by the total number of trials, is strongly dependent on the learning rates. For large values of q_-, the switching probability is high, but a large value of q_+ reduces the switching probability. ***D***, The range of parameters for which the model shows an adequate matching behavior is plotted in red, that is, when $l_{tot}/\lambda_{tot} > 0.74$ [this quantity is ~0.72 for the monkeys (Sugrue et al., 2004a)] and the deviation from matching is <0.1. For each set of model parameters, all average values are obtained from 1000 simulated sessions of the experiment (see Materials and Methods). The length of each block is set to 200 trials and $\sigma = 5\%$.

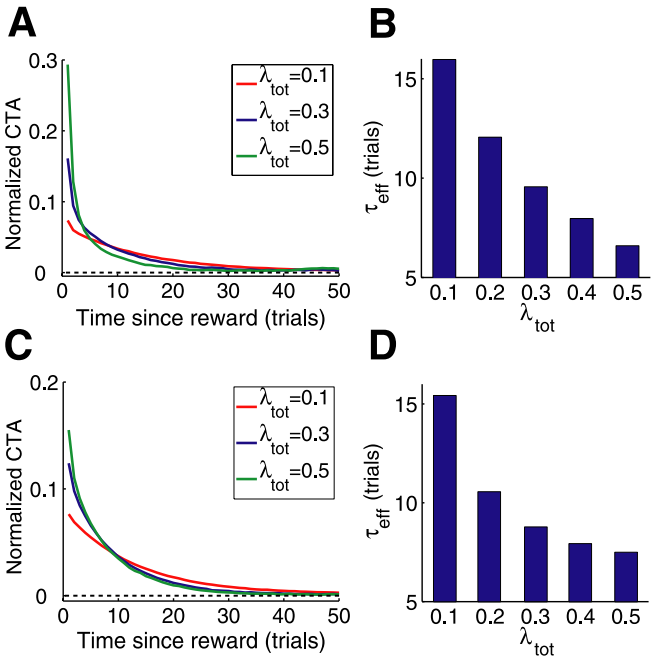

Figure 12. Time integration of past rewards and its dependence on the overall baiting probability. ***A***, The choice-triggered average of rewards, extracted from the model choice behavior, is plotted for three different values of overall baiting probability (for a fixed set of the model parameters; $q_+ = 0.15$, $q_- = 0.05$, $\sigma = 25\%$). As the overall baiting probability, λ_{tot}, increases, recent rewards have a stronger effect, but the effect of past rewards decays more quickly. ***B***, The effective time constant of CTA, defined as the weighted sum of the two extracted time constants, is plotted as a function of the overall baiting probability. As the total reward rate increases, the effective time constant decreases. ***C***, For a different set of model parameters ($q_+ = 0.06$; $q_- = 0.06$; $\sigma = 5\%$), CTA extracted from the model choice behavior is plotted for three different values of overall baiting probability. ***D***, The effective time constant of CTA extracted from the choice behavior of the model for the same set of parameters as in ***C***.

optimal performance. There is an intermediate range of learning rates that results in a large reward rate and also reasonable matching behavior. We define a set of the model parameters as suitable, if for such a set of parameters the performance of the model is >74% and the deviation from matching is <0.1. These sets of parameters are shown in Figure 11*D* in red. Note that suitable behavior can be achieved for a wide range of the learning rates, so that in order for the model to perform the matching task, fine-tuning of the learning parameters is not necessary.

Dependence of the choice on the past rewards

To quantify the dependence of choice in each trial on the past history of reward, we follow Sugrue et al. (2004a) (Corrado et al., 2005) to calculate what they termed as the "choice-triggered-average of rewards (CTA)." This quantity measures how choice in the current trial is influenced by the harvested rewards in the past trials. Here, we are mainly interested in how the form of CTA, extracted from the model choice behavior, is influenced by external factors such as the length of blocks in which the baiting reward rate is constant, and the overall baiting probability in the experiment.

In general, the form of CTA in our model can vary, depending on the learning rates and the noise level σ. Interestingly, we find that the shape of CTA is independent of the length of the blocks, consistent with the observation of Sugrue et al. (2004a) that CTA is the same no matter which part of the block is used for analysis. To compare our results with the Sugrue et al. (2004a) experiment, we choose a set of the model parameters that results in a CTA close to the CTA extracted from one of the monkeys in their experiment. Using the same parameters, we allow the model to play the same task with different overall baiting probabilities (λ_{tot}).

The CTA for three different values of overall baiting probabilities is shown in Figure 12*A*. When the overall baiting probability is high, the dependence of the choice on the recent harvested rewards is stronger, but this dependence decreases more rapidly with time (in unit of trial).

In contrast, when the overall baiting probability is low, the choice is influenced by past rewards over a longer time. To quantify this dependence, the CTA obtained from the behavior of the model is fitted with the sum of two exponentials as follows:

$$\mathrm{CTA}(t) = \omega N_s e^{-\frac{t}{\tau_s}} + (1-\omega) N_l e^{-\frac{t}{\tau_l}}, \qquad (8)$$

where t is the trial number, N_s and N_l are the normalization factors for each exponential, ω is the weighting factor, and τ_s and τ_l are the short and long time constants, respectively. The sum of two exponentials provides a good fit for the CTA extracted from the monkeys' data (Corrado et al., 2005). The results of fitting show that, when the overall baiting probability increases, the longer time constant, τ_l, and its relative weight, $1-\omega$, decrease (data not shown). To simplify the comparison, we define an effective time constant, $\tau_{eff} = \omega\tau_s + (1-\omega)\tau_l$. The effective time constant displays the approximate timescale over which the inte-

gration of past rewards is performed. As shown in Figure 12*B*, the effective time constant decreases as the overall baiting probability increases.

For the set of model parameters used in most of the simulations ($q_+ = 0.06$; $q_- = 0.06$; $\sigma = 5\%$), the extracted form of CTA is different from the experimental observation (Fig. 12*C*,*D*), namely, in this case the relative weight of the shorter time constant is small. Additional study shows that generally a more biexponential CTA can be achieved if the choice behavior of the model is more random (larger value of σ), which results in more undermatching. Although generally the quality of matching in monkeys is poorer than the model, these results may indicate that there are other factors that influence the monkey's choice behavior, such as past choices of the monkey, which are not included in the present model.

Discussion

Recent neurophysiological studies of nonhuman primates performing probabilistic decision making tasks showed that single-cell activities in certain brain areas are modulated by the subjective values of choice options (Kawagoe et al., 1998; Leon and Shadlen, 1999; Platt and Glimcher, 1999; Lauwereyns et al., 2002; Montague and Berns, 2002; Shidara and Richmond, 2002; Barraclough et al., 2004; Sugrue et al., 2004a, 2005; Samejima et al., 2005; Hikosaka et al., 2006). In this study, we addressed the question of how these subjective values may be computed mechanistically and used to generate choice behavior. We showed that plastic synapses that undergo stochastic reward-dependent modification can integrate past rewards within a finite time window. This is because synapses have a limited number of discrete states, so reward history in the remote past is forgotten, resulting in a finite integration time. The strengths of these synapses influence the choice behavior in any given trial, and in turn they are modified depending on the choice made and the resulting outcome in that trial. This two-way interplay between synaptic plasticity and the decision process gives rise to trial-by-trial adaptive choice behavior in a dynamic and stochastic environment. In this work, we applied our model to a matching task paradigm and showed that, under certain conditions, the average strength of plastic synapses onto a choice-selective neural population is equal to the return from that choice. Therefore, we propose that subjective values can be computed dynamically, in the form of return, at the synaptic level. Furthermore, decision neurons that receive inputs from these plastic synapses exhibit graded levels of activity. The latter is modulated by the choice of the network and the difference in the average synaptic inputs to the competing neural populations. In this way, the subjective values computed at a synaptic level become observable in the spike firing neural activities of a decision-making network.

Learning rule

Learning that depends on reinforcement feedback signals (Sutton and Barto, 1998) is believed to underlie many adaptive behaviors in a natural environment. Evidence suggests that dopamine in the brain acts as a common currency for a reward signal (Schultz, 2000, 2006), and modulates synaptic plasticity (Jay, 2003). For instance, at the corticostriatal synapses onto the medium spiny projection neurons in striatum, long-term plasticity depends on stimulations of dopamine neurons (Reynolds et al., 2001; Reynolds and Wickens, 2002) (but see Fino et al., 2005). Based on these observations, Reynolds and Wickens (2002) proposed a three-factor synaptic plasticity rule, in which synaptic modifications depend on presynaptic and postsynaptic neural activities as well as a dopamine signal. Other studies indicate that, in the rat prefrontal cortex, the induction of long-term depression and long-term potentiation at glutamatergic synapses is modulated by dopamine (Otani et al., 2003), and that D_1 receptors play a key role in such bidirectional modulation of plasticity (Huang et al., 2004).

In this work, we sought to implement such a three-factor learning rule in a biophysically plausible manner (Fusi et al., 2005a). Our learning rule is Hebbian and depends on the coactivation of presynaptic and postsynaptic neurons. Moreover, individual synapses have two discrete states (Petersen et al., 1998; O'Connor et al., 2005), and plasticity occurs as a stochastic process (Amit and Fusi, 1994; Fusi, 2002). The fact that synapses are bounded is important, because in this way the number of available synaptic states are limited and this enables the model to forget the past outcomes naturally. The stochastic nature of plasticity implies that modifications occur at every single trial, which form the basis of adaptive decision process; yet the average synaptic changes take place over many trials (determined by the learning rates), over a timescale compatible with that of experimental protocols for the induction of long-term synaptic potentiation or depression (Malenka and Nicoll, 1999; Bi and Poo, 2001). Finally, the direction of modification (potentiation versus depression) is reversed by the presence/absence of reward (see below for variants of this learning rule). Seung (2003) also considered a reward-dependent synaptic plasticity rule and proposed an algorithm based on a reward signal that modulates the probability of stochastic release of transmitters. By design, Seung's algorithm maximizes rewards under general conditions, whereas this is not guaranteed with our model. In contrast, Seung's model performs better if an all-or-none reward signal, like dopamine, is delivered every time a presynaptic spike is fired; how this may be accomplished biologically remains unclear. Delivery of reward signal with a delay, requires an integration of eligibility trace over a timescale of seconds during the decision process and makes the learning process very slow. In contrast, in our model, plasticity occurs only once in a trial, rather than continuously, at the time of potential reward delivery.

We focused on a specific learning rule in which plasticity takes place only when the postsynaptic neurons fire spikes at a high rate, in other words, only for plastic synapses projecting to the neural population that has won the competition in a given trial. As we have shown, our model exhibits satisfactory matching behavior comparable with observations in the experiment of Sugrue et al. (2004a), robustly for a wide range of learning rates for potentiation and depression, respectively. The only condition is that the potentiation rate should be similar to or larger than the depression rate. We also explored other variants of learning rules. For example, plasticity could take place without requiring high firing activity of postsynaptic neurons. Thus, in each trial both sets of plastic synapses onto the two competing neural populations are modified. In this situation, the synaptic strength is not a function of return. Instead, it is a function of income (for a more detailed description of such a model and its behavior, see supplemental material, available at www.jneurosci.org).

Moreover, if the decision criterion is given by the fractional income [i.e., $P_A = I_A/(I_A + I_B)$ (like in the model of Sugrue et al., 2004a)] instead of a sigmoid function of the differential income, the choice behavior may become unstable. The instability happens in the income-based model for the following reason: if one of the targets is consecutively chosen, although few rewards are obtained, the income on the chosen target fluctuates around some level while the income on the unchosen target goes to zero. This further decreases the probability of selecting the unchosen

target and results in repeated selection of one of the targets. In contrast, the return changes for the selected target only, hence such instability does not occur in a return-based decision model.

If the rule is such that, in a rewarded trial, not only synapses are strengthened with high postsynaptic activity (of neurons selective to the chosen target), but also weakened with low postsynaptic activity (of neurons selective for the unchosen target), then the model tends to select the target with a higher baiting probability excessively, a phenomenon called "overmatching." One could argue that overmatching may be avoided using a learning rule according to which, in an unrewarded trial, potentiation occurs with low postsynaptic firing rates (for those synapses projecting to neurons selective for the unchosen target). This plasticity rule seems biophysically implausible. Moreover, this rule typically leads to more undermatching than the rule presented in this paper does, and is thus functionally undesirable (large deviation from matching). The general conclusion is that the most suitable and robust learning rule for the matching task is the one in which only plastic synapses related to the selected choice undergo plasticity, and this rule is qualitatively in agreement with the available experimental evidence for reward-dependent synaptic plasticity.

Mechanisms of matching behavior

Although matching behavior has been observed in many different experimental paradigms, how it is achieved by a local (trial-to-trial) decision process is still not fully understood. In one scenario, Gallistel et al. (2001) proposed that a local mechanism based on ideal detectors of changes in reward rates can account for matching behavior. Another proposal relies on the idea that matching is a manifestation of reward maximization (Staddon and Motheral, 1978; Baum, 1981). A third theory, called "melioration" (Herrnstein and Vaughan, 1980; Williams, 1988; Herrnstein and Prelec, 1991), posits a decision dynamics in which the subject chooses the behavioral alternative that provides the higher local reinforcement rate (or return) at that time. In the special case of the concurrent variable-interval schedule, this local mechanism produces global matching behavior, because an increase (decrease) in the selection of one option decreases (increases) the return from that option. Therefore, an equilibrium is reached when the returns from the two alternatives are equal. However, in general, melioration can result in a behavior different from matching. In fact, in experiments in which melioration, matching, and maximization give different predictions, behavioral data were consistent with the melioration theory (Vaughan, 1981).

Two key issues have been left unresolved in the melioration theory. First, as stated by Williams (1988): "the most fundamental problem faced by melioration theory is the specification of the method by which local reinforcement rates are calculated." Second, melioration has often been taken to mean "choose the option with the highest return among all possible alternatives." Although this deterministic rule yields matching as the steady state, the stability of that behavior is not guaranteed. The neural model reported in this paper sheds insights into both issues.

Our model proposes a neurobiological implementation of melioration. First, in our model, a local estimate of return (or a function of return) on each choice is computed by synapses that undergo reward-gated stochastic plasticity. We found that there is a tradeoff between the accuracy of the estimated return and the flexibility of choice behavior. If the learning rates are low, synapses can integrate rewards over a long period of time and the estimation of return would be accurate. However, this means that the system cannot adapt quickly when the reward schedule changes frequently in an uncertain environment. In contrast, if the learning rates are higher, the behavior is more flexible, but the reward integration is more local in time and the estimated return is noisier. This raises the interesting question of whether learning rates themselves should be plastic (meta-learning) and adjustable according to behavioral demands (Doya, 2002; Schweighofer and Doya, 2003). We intend to address this question elsewhere.

Second, in our model, decision making is not deterministic even if the returns of options are known. Instead, we showed that, in a recurrent circuit of spiking neurons, the choice probability is a softmax (sigmoid) function of the difference in the returns (coded by the strengths of synapses to the two competing neural populations). This stochasticity is attributable to irregular spike discharges, a characteristic feature of cortical neurons (Softky and Koch, 1993; Shadlen and Newsome, 1994; van Vreeswijk and Sompolinsky, 1996; Compte et al., 2003). The more variable the neuronal firing activity, the less steep is the softmax function (with a larger σ). Thus, too much noise would mean a very graded softmax decision criterion; the choice behavior would be essentially random and far from matching. In contrast, with negligible noise (small σ value), the system has a tendency to only choose the target with a higher return, and this may result in instability of the choice behavior. To avoid this kind of instability, the network should be able to make decisions probabilistically. We also showed that probabilistic decision making imposes a limit on how close matching can be achieved. Therefore, our model provides a possible explanation, in terms of neural network constraints, for undermatching, a phenomenon widely observed across different species (Baum, 1974, 1979; Davison and Baum, 2000; Anderson et al., 2002; Sugrue et al., 2004a; Lau and Glimcher, 2005b).

Our model has similarities to, as well as differences with, other recently proposed models for matching behavior. For example, the model of Sugrue et al. (2004a) assumes that local incomes on two options are computed by a leaky integrator and then these quantities are used to compute the local fractional income. If the instantaneous probability of choice is equal to the local fractional income, the model obeys the matching law locally. This model provides a good account of monkeys' behavioral data but leaves open mechanistic questions, such as how integration over the income is done, how the time constant for the integration is determined in the circuit, how local fractional income is calculated (which requires two additional computations, addition and division), and how it can be translated to choice probability. In a revision of this model, Corrado et al. (2005) replaced the decision rule according to fractional income, by a softmax function of the difference in local incomes. Our model represents a biophysical instantiation of that scenario, except that it uses return rather than income.

Furthermore, our work suggests a synaptic mechanism for the linear filter (called CTA) that has been deduced from behavioral data and hypothesized to subserve reward integration (Sugrue et al., 2004a; Corrado et al., 2005; Lau and Glimcher, 2005b). In our model, the time constants of CTA are determined by the model parameters. Hence, the experimentally observed CTAs can be associated with biophysical quantities like potentiation and depression rates at the synaptic level. Importantly, we showed that the overall baiting probability influences the form of CTA, so that the integration times are stretched or contracted depending on how abundant rewards are in the environment. Specifically, we showed that, when the overall baiting probability is lower (higher), because synaptic modifications depend on reward frequency, the effective integration time becomes larger (smaller) so that rewards are integrated over a longer (shorter) timescale,

which makes sense functionally. This prediction of our model can be readily tested by varying the overall reward rate in matching task experiments. Finally, our model semiquantitatively reproduces electrophysiological data recorded from behaving monkeys, whereas previous models (Sugrue et al., 2004a; Corrado et al., 2005; Lau and Glimcher, 2005b) were mostly focused on behavioral data.

LIP neurons: representation of decision or value?

Neurons in the LIP area of the posterior parietal cortex show activity that is believed to be important for guiding saccadic eye movements. In experimental studies of the matching task, it has been shown that these neurons carry information about the impending movements and the subjective value of those movements. Platt and Glimcher (1999) showed that the activity of some LIP neurons is modulated by the gain of the choice into the response field (RF) of the neuron. Sugrue et al. (2004a) showed that activity of LIP neurons is modulated by the impending choice and the fractional income for that choice. Importantly in a given trial, the activity of a neuron is higher if the monkey's choice is into the RF of the neuron than if the monkey's choice is out of the RF of the neuron. In addition, for a fixed choice, the activity of a neuron is higher when the local fractional income of the RF target of the neuron is higher. In a later paper, Corrado et al. (2005) showed that, in fact, their neural data were better described as being correlated with the difference between the local incomes from the two targets, rather than with the fractional income.

Although the graded activity of neurons in area LIP carries information about the value for each choice, there is evidence that the valuation is not computed in LIP. Indeed, LIP neurons are spatially selective and not color selective, whereas in this task the rewarding value of each target is coded by the target color and not by its location. Moreover, the time course of neural activity becomes differentiated according to the income level of a given target, at least 100–200 ms after the stimulus onset, indicating that the value-related information originates from somewhere else.

Our working hypothesis is that the primary role of LIP neurons is to generate a decision about saccadic eye movement, based on an integration of two types of inputs: spatial target and its rewarding value. According to this view, for a given neuron (and a selected target), the firing activity depends parametrically on its overall input. Thus, graded activity emerges whenever the trials are sorted according to different levels of the overall input of the neuron, regardless of whether it is a sensory or reward signal, or a combination of both. This interpretation is consistent with the observation that graded LIP neural activities, similar to those in the study by Sugrue et al. (2004a), were also found in a visual motion direction discrimination task in which the differentiating factor is motion coherence (sensory information) rather than rewarding value (Shadlen and Newsome, 2001; Roitman and Shadlen, 2002). It is also in line with the finding that, when the two choices about motion direction are associated with different amounts of reward, the subject's psychometric function is shifted in such a way as if reward magnitude provided an extra signal that is additive to the sensory information about the motion direction (Rorie and Newsome, 2004).

Here, we showed that a model based on this idea reproduces graded neural activities observed in LIP. In our model, neurons are responsible for making decisions, hence the spiking activities are correlated with and give rise to choices. In addition, because input synapses to these neurons encode reward history, neuronal firing rates naturally reflect the target rewarding values. Because the choice probability and returns are directly related to each other (via the softmax function), it is impossible to dissociate the two. Similarly, in a physiological experiment, correlations between neural activity and the subjective value of choice options do not necessarily imply that the recorded neurons (like LIP cells) are primarily involved with valuation rather than decision making.

It is worth noting that, as our results here demonstrate, graded activities are compatible with the attractor dynamics of our model. Indeed, although an attractor network displays stable activity states in the absence of direct stimulation (e.g., during a delay period of working memory), it is readily configurable by external inputs and can depend on input strength in a graded manner. Moreover, the other aspect of the observed graded LIP activity, namely the divergence over time of firing activities corresponding to two alternative choices, is explained in our model by effective mutual inhibition between the two selective populations. Because of this competition, when the firing rate of one selective population is high, that of the other selective population goes down. This is similar to what has been observed in the experiment of Sugrue et al. (2004a). In their experiment, if the left choice has the highest local fractional income and it is selected, then neurons with RF on the left target have the highest level of activity and neurons with RF on the right target have the lowest level of activity. If we assume that neurons with RF on the left and right target belong to two different pools of neurons in LIP, then the most plausible explanation for the above observation is the existence of competition between these two pools of neurons. This, again, is consistent with the suggestion that the LIP neurons behave like decision makers, or have an important role in the decision-making processes.

In this paper, we focus on a microcircuit model endowed with synaptic plasticity. In all likelihood, this model will need to be expanded, and the following alternative scenarios should be considered in future studies. First, if integration of past rewards is performed by plastic synapses, it is an open question as to the precise locus (or loci) of such plasticity. In addition to LIP, other candidate sites include corticostriatal synapses in basal ganglia, or synaptic connections in the orbitofrontal cortex (Schultz, 2000). For example, a new study showed that postsaccadic activity of caudate neurons encodes the preceding saccade and/or reward delivery in a matching task experiment (Lau and Glimcher, 2005a). Therefore, a large-scale network with multiple interacting brain areas should be investigated. Secondly, it is conceivable that past rewards can be integrated by cellular mechanisms in single neurons, instead of plastic synapses. There is evidence that neural activity in the dorsolateral prefrontal cortex (Barraclough et al., 2004) and orbitofrontal cortex (Sugrue et al., 2004b) is modulated by reward signals across trials. However, it remains an open question whether these reward-modulated neural activities are generated by a cellular or synaptic mechanism. Additional experimental and computational work will shed light on this fundamental question about the neurobiological basis of choice behavior.

References

Amit DJ, Fusi S (1994) Dynamic learning in neural networks with material synapses. Neural Comput 6:957–982.

Anderson KG, Velkey AJ, Woolverton WL (2002) The generalized matching law as a predictor of choice between cocaine and food in rhesus monkeys. Psychopharmacology (Berl) 163:319–326.

Barraclough DJ, Conroy ML, Lee D (2004) Prefrontal cortex and decision making in a mixed-strategy game. Nat Neurosci 7:404–410.

Baum WM (1974) 2 Types of deviation from matching law-bias and undermatching. J Exp Anal Behav 22:231–242.

Baum WM (1979) Matching, undermatching, and overmatching in studies of choice. J Exp Anal Behav 32:269–281.

Baum WM (1981) Optimization and the matching law as accounts of instrumental behavior. J Exp Anal Behav 36:387–403.
Bi G, Poo M (2001) Synaptic modification by correlated activity: Hebb's postulate revisited. Annu Rev Neurosci 24:139–166.
Brunel N, Wang X-J (2001) Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. J Comput Neurosci 11:63–85.
Brunel N, Carusi F, Fusi S (1998) Slow stochastic hebbian learning of classes of stimuli in a recurrent neural network. Network 9:123–152.
Compte A, Constantinidis C, Tegner J, Raghavachari S, Chafee MV, Goldman-Rakic PS, Wang X-J (2003) Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. J Neurophysiol 90:3441–3454.
Corrado GS, Sugrue LP, Seung HS, Newsome WT (2005) Linear-nonlinear-Poisson models of primate choice dynamics. J Exp Anal Behav 84, 581–617.
Davison M, Baum WM (2000) Choice in a variable environment: every reinforcer counts. J Exp Anal Behav 74:1–24.
de Villiers PA, Herrnstein RJ (1976) Toward a law of response strength. Psychol Bull 83:1131–1153.
Dorris MC, Glimcher PW (2004) Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron 44:365–378.
Doya K (2002) Metalearning and neuromodulation. Neural Netw 15:495–506.
Fino E, Glowinski J, Venance L (2005) Bidirectional activity-dependent plasticity at corticostriatal synapses. J Neurosci 25:11279–11287.
Fusi S (2002) Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. Biol Cybern 87:459–470.
Fusi S, Asaad WF, Miller EK, Wang X-J (2005a) A microcircuit model of arbitrary sensori-motor mapping: learning and forgetting on multiple timescales. Soc Neurosci Abstr 31:813.10.
Fusi S, Drew PJ, Abbott LF (2005b) Cascade models of synaptically stored memories. Neuron 45:599–611.
Gallistel CR (1994) Foraging for brain stimulation: toward a neurobiology of computation. Cognition 50:151–170.
Gallistel CR, Mark TA, King AP, Latham PE (2001) The rat approximates an ideal detector of changes in rates of reward: implications for the law of effect. J Exp Psychol Anim Behav Process 27:354–372.
Herrnstein RJ (1961) Relative and absolute strength of response as a function of frequency of reinforcement. J Exp Anal Behav 4:267–272.
Herrnstein RJ, Prelec D (1991) Melioration: a theory of distributed choice. J Econ Perspect 5:137–156.
Herrnstein RJ, Vaughan WJ (1980) Melioration and behavioral allocation. In: Limits to action: the allocation of individual behavior (Staddon JER, ed), pp 143–176. New York: Academic.
Herrnstein RJ, Rachlin H, Laibson DI (1997) The matching law: papers in psychology and economics. Cambridge, MA: Harvard UP.
Heyman GM (1979) A Markov model description of changeover probabilities on concurrent variable-interval schedules. J Exp Anal Behav 31:41–51.
Heyman GM, Luce D (1979) Operant matching is not a logical consequences of maximizing reinforcement rate. Learn Behav 7:133–140.
Hikosaka O, Nakamura K, Nakahara H (2006) Basal ganglia orient eyes to reward. J Neurophysiol 95:567–584.
Houston AI, Sumida BH (1987) Learning rules, matching and frequency dependence. J Theor Biol 126:289–308.
Huang Y-Y, Simpson E, Kellendonk C, Kandel ER (2004) Genetic evidence for the bidirectional modulation of synaptic plasticity in the prefrontal cortex by D1 receptors. Proc Natl Acad Sci USA 101:3236–3241.
Jay TM (2003) Dopamine: a potential substrate for synaptic plasticity and memory mechanisms. Prog Neurobiol 69:375–390.
Kawagoe R, Takikawa Y, Hikosaka O (1998) Expectation of reward modulates cognitive signals in the basal ganglia. Nat Neurosci 1:411–416.
Lau B, Glimcher PW (2005a) Caudate neurons encode both saccade and reward information in a free-choice task. Soc Neurosci Abstr 31:400.14.
Lau B, Glimcher PW (2005b) Dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav 84, 555–579.
Lauwereyns J, Watanabe K, Coe B, Hikosaka O (2002) A neural correlate of response bias in monkey caudate nucleus. Nature 418:413–417.
Leon MI, Shadlen MN (1999) Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. Neuron 24:415–425.
Malenka RC, Nicoll RA (1999) Long-term potentiation—a decade of progress? Science 285:1870–1874.
Montague PR, Berns GS (2002) Neural economics and the biological substrates of valuation. Neuron 36:265–284.
O'Connor DH, Wittenberg GM, Wang SS-H (2005) Graded bidirectional synaptic plasticity is composed of switch-like unitary events. Proc Natl Acad Sci USA 102:9679–9684.
Otani S, Daniel H, Roisin M-P, Crepel F (2003) Dopaminergic modulation of long-term synaptic plasticity in rat prefrontal neurons. Cereb Cortex 13:1251–1256.
Petersen CC, Malenka RC, Nicoll RA, Hopfield JJ (1998) All-or-none potentiation at CA3-CA1 synapses. Proc Natl Acad Sci USA 95:4732–4737.
Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. Nature 400:233–238.
Reynolds JN, Wickens JR (2002) Dopamine-dependent plasticity of corticostriatal synapses. Neural Netw 15:507–521.
Reynolds JN, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. Nature 413:67–70.
Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J Neurosci 22:9475–9489.
Rorie AE, Newsome WT (2004) The role of area LIP in a direction discrimination task with multiple reward contingencies. Soc Neurosci Abstr 30:20.11.
Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337–1340.
Schmolesky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AG (1998) Signal timing across the macaque visual system. J Neurophysiol 79:3272–3278.
Schultz W (2000) Multiple reward signals in the brain. Nat Rev Neurosci 1:199–207.
Schultz W (2006) Behavioral theories and the neurophysiology of reward. Annu Rev Psychol 57:87–115.
Schweighofer N, Doya K (2003) Meta-learning in reinforcement learning. Neural Netw 16:5–9.
Seung HS (2003) Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. Neuron 40:1063–1073.
Shadlen MN, Newsome WT (1994) Noise, neural codes and cortical organization. Curr Opin Neurobiol 4:569–579.
Shadlen MN, Newsome WT (1996) Motion perception: seeing and deciding. Proc Natl Acad Sci USA 93:628–633.
Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86:1916–1936.
Shidara M, Richmond BJ (2002) Anterior cingulate: single neuronal signals related to degree of reward expectancy. Science 296:1709–1711.
Softky WR, Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. J Neurosci 13:334–350.
Soltani A, Wang X-J (2004) Exploring the neural basis of the matching law in choice behavior: a cortical network model with reward-gated learning. Soc Neurosci Abstr 30:668.14.
Staddon JER, Motheral S (1978) On matching and maximization in operant choice experiments. Psychol Rev 85:436–444.
Sugrue LP, Corrado GC, Newsome WT (2004a) Matching behavior and representation of value in parietal cortex. Science 304:1782–1787.
Sugrue LP, Corrado GC, Newsome WT (2004b) Neural correlates of value in orbitofrontal cortex of the rhesus monkey. Soc Neurosci Abstr 30:671.8.
Sugrue LP, Corrado GS, Newsome WT (2005) Choosing the greater of two goods: neural currencies for valuation and decision making. Nat Rev Neurosci 6:363–375.
Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT.
van Vreeswijk C, Sompolinsky H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science 274:1724–1726.
Vaughan W (1981) Melioration, matching, and maximization. J Exp Anal Behav 36:141–149.
Wang X-J (2002) Probabilistic decision making by slow reverberation in cortical circuits. Neuron 36:955–968.
Williams BA (1988) Reinforcement, choice, and response strength. In: Steven's handbook of experimental psychology, Ed 2, Vol 2 (Atkison RC, Herrnstein RJ, Lindzey G, Luce RD, eds), pp 167–244. New York: Wiley.