SUPPLEMENTARY MATERIALS

1 A schematic model for remapping from color to location

In the experiment of Sugrue et al. (2004) the rewarding value (baiting probability) of each target is determined by the target color (red and green) and not its location. Moreover, the location of red and green targets were assigned randomly in each trial. On the other hand, LIP neurons are selective to target location. So in order for LIP neurons to receive the information about the rewarding value of each target, a remapping from color to location should take place in each trial. Here we describe the scheme of a model endowed with a remapping circuit (Fig. 1A). Furthermore, we show how this model can be reduced to the simple model which we have used in the paper (Fig. 1B).

The detailed model consists of three layers of neurons (Fig. 1A). The first layer consists of two populations of neurons which are selective to target color and receive sensory inputs through some plastic synapses. The second (intermediate) layer of neurons is responsible for the remapping from color to location. It consists of two color selective populations of neurons, each one consisting of two subpopulations which receive input about the location of each target. Hence, there are four neural subpopulations selective for combinations of the target color and position. The two populations in the intermediate layer effectively inhibit each other through an inhibitory population of inter-neurons. The third layer of neurons is the spatially selective decision-making network which has been described in the paper.

Upon the presentation of the two targets, neurons in the first layer receive color specific sensory inputs through plastic synapses which undergo reward-dependent learning. As a result, the activity of these neurons is modulated by the synaptic strengths of the input plastic synapses. The activity of these neurons consequently influences the activity of neurons in the intermediate layer. In addition to the inputs from the first layer, neurons in the intermediate layer receive inputs about the location of each target. For example, if the green target appears on the left side (and the red target on the right side), then only the red-right and green-left subpopulations in the interme-

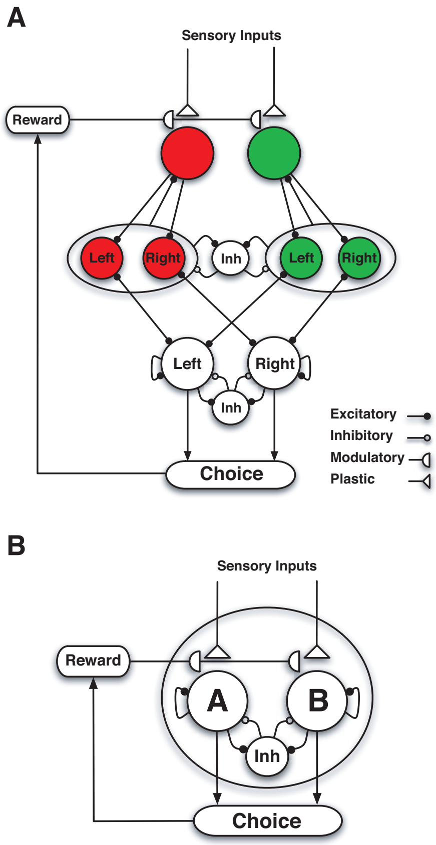

Figure 1: Scheme of the detailed (A) and simplified (B) models. (A)The detailed model contains three layers of neurons. The first layer consists of two populations of neurons which are selective to the red and green targets. These neurons receive sensory information through some plastic synapses which undergo reward-dependent plasticity. The second layer consists of two populations selective to target color and each of these populations contains two subpopulations which receive input about the location of each target. The two population in the intermediate layer effectively inhibit each other through an inhibitory population of inter-neurons. The third layer is the decision-making network which has been described in the paper. There is reciprocal connections between neurons in the first and second layers, and between the second and third layers. (B) The simplified model contains two populations of neurons which are selective to the two targets and receive direct sensory inputs through some plastic synapses.

diate layer are activated. These neurons project to corresponding spatially selective decision-making neurons (Fig. 1A). In this way, the rewarding information of each target, stored in the input plastic synapses to the first layer, are correctly transferred to the left and right populations.

At the end of each trial, the decision is made when one of the selective populations (say, Right) in the decision-making network reaches a high level of activity. The high level of activity in the selective population Right in the decision-making network increase the level of activity in one of the already active subpopulations, namely red-right, in the intermediate layer (through reciprocal connection). This in turn suppresses the activity in the two subpopulations selective to the other color, green, because of mutual inhibition in the intermediate layer. Backprojection from the intermediate to the first layer leads to a large increase in the firing rates in one of the two color-selective neural populations (green). As a result, at the end of each trial only one of the populations in the first layer becomes highly active. After the choice is made and at the time of reward delivery, synapses projecting to the chosen color target which has a high level of activity are modified.

For the sake of computational efficiency and due to lack of enough experimental information on the remapping from color to location, we use a simplified version of the detailed model (Fig. 1B). The simplified model consists of a general decision-making network which includes two populations of neurons selective to target options, and receive inputs via plastic synapses. In this way we do not specifically address the important issue of color-to-location remapping in the present work.

2 A variant of learning rule leads to an income-based decision model

Here we consider a less biophysically plausible learning rule in which both sets of plastic synapses undergo modification in each trial. One possible scenario is that in the rewarded trials synapses projecting to the chosen population are potentiated (with learning rate q_r) and synapses projecting to the unchosen population are depressed (with the same learning rate q_r). In the unrewarded trials both sets of plastic synapses are depressed (with learning rate q_n). We show that for this learning rule the steady state of synaptic strengths of the two sets of plastic synapses are approximately equal to the income from the two choices (This approximation holds while the learning is slow, i.e. when q_r and q_n are small).

The probability of obtaining a reward on target i ($i = A$ or B) is equal to income from the same target ,I_i. If the probability of choosing target i is P_i, then the average change in the synaptic strength, c_i, in each trial is given by

$$\Delta c_i = q_r(1 - c_i)I_i - q_r c_i I_k - q_n c_i(P_i - I_i) - q_n c_i(1 - P_i - I_k) \quad (k \neq i)$$

The first term is the change due to potentiation in a rewarded trial (which occurs with the probability of I_i) The second term is the change due to depression in a trial in which target k ($k \neq i$) is rewarded (which occurs with the probability of I_k) . The third and forth terms are changes due to depression in trials with no reward (which occurs with the probability of $P_i - I_i$ for target i and $1 - P_i - I_k$ for target k).

In the steady state, Δc_i should be zero which results in

$$c_i^{ss} = \frac{q_r I_i}{(q_r - q_n)(I_i - I_k) + q_n} \tag{1}$$

In the special case for which $q_r = q_n$, the steady state of c_i is equal to the income from the same target, I_i. If q_r and q_n are different, c_i is roughly a linear function of the income, $c_i^{ss} \simeq (q_r/q_n)I_i$, as long as $|q_r - q_n|(I_i - I_k)$ is much smaller than q_n. The latter inequality generally holds when the income is significantly smaller than 1 and q_r is not much larger than q_n. These results are confirmed by simulations of the matching task experiment (Fig. 2). As shown in Fig. 2 the average synaptic strengths is a linear function of the income in each block of the experiment.

The income-based model described here is fairly similar to the model used in Corrado et al. (2005) except that in our model the integration of past rewards is performed by the plastic synapses, while it was done by a presumed filter in Corrado et al. (2005).

The income-based model is not as robust as the return-based model presented in the paper. To show this point we plot the average measures of the choice behavior of the income-based model in the matching task experiment (Fig. 3). First of all, the performance of this model changes more drastically as a function of the model parameters (Fig. 3A). Secondly, the deviation from matching is comparable to the return-based model but this is mostly achieved by long stays on each choice (see Fig. 3C). As shown in Fig. 3D

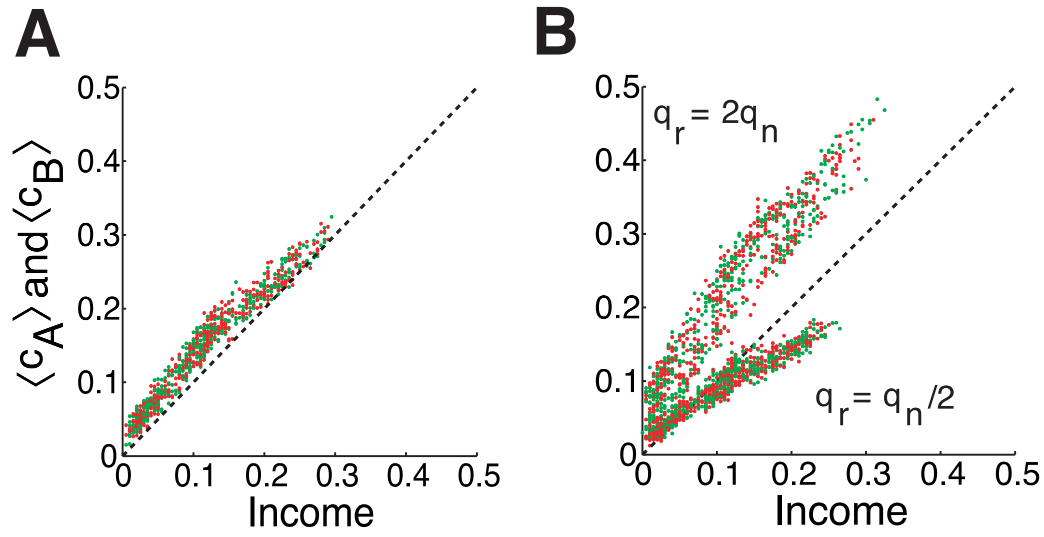

Figure 2: Plastic synapses approximately compute the income from each choice. Block-averaged synaptic strengths are plotted vs. the obtained incomes in the same block. Two colors represent two different choices (red for A and green for B). (A) The average synaptic strength is equal to the income from each choice, when the two learning rates are equal ($q_r = q_n = 0.06$). (B) If the learning rate in rewarded trials is larger than in unrewarded trials ($q_r = 0.06, q_n = 0.03$), the average synaptic strength is larger than the income (upper points); whereas, if the learning rate in rewarded trials is smaller than in unrewarded trials ($q_r = 0.03, q_n = 0.06$), then the synaptic strength is smaller than the income (lower points). These data points are obtained from 25 simulated sessions of the matching task in which all possible baiting probability ratios are used (see Methods) and the length of each block is set to 200 trials. For all simulations σ is set to 10%.

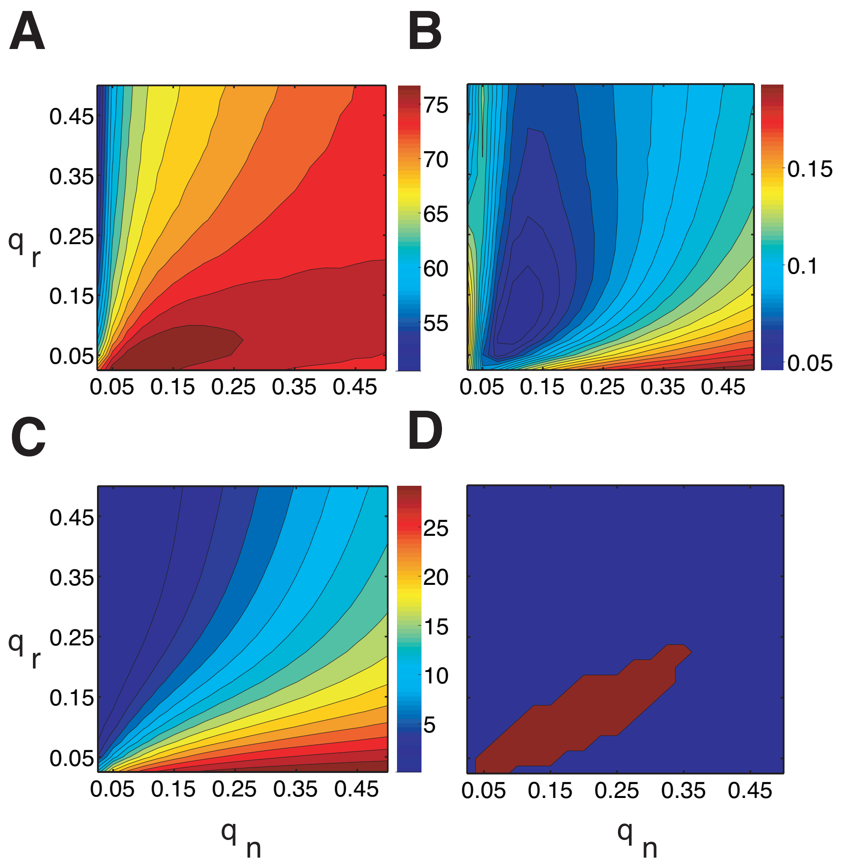

Figure 3: Model shows matching behavior over a narrow range of parameters. (A) The model's performance I_{tot}/λ_{tot}, defined as the ratio of the average reward rate to the overall baiting probability, changes significantly with learning rates. (B) The 'deviation from matching', computed as the average of absolute difference between choice and reward fractions on each block, is small over a wide range of learning rates. (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_n, the switching probability is high but a large value of q_r 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 $I_{tot}/\lambda_{tot} > 0.74$, and the deviation from matching is less than 0.1. For each set of model parameters all average values are obtained from 1000 simulated sessions of the experiment. The length of each block is set to 200 trials and $\sigma = 5\%$.

sets of model parameters for which a reasonable matching is achieved are restricted to a much smaller area than the return-based model presented in the paper. These results show that a less biophysically plausible income-based model requires more fine-tuning than the return-based model.

References

Corrado, G. S., L. P. Sugrue, H. S. Seung, and W. T. Newsome (2005). Linear-Nonlinear-Poisson models of primate choice dynamics. *J Exp Anal Behav 84*, 581–617.

Sugrue, L. P., G. C. Corrado, and W. T. Newsome (2004). Matching behavior and representation of value in parietal cortex. *Science 304*, 1782–1787.