



Reward-biased probabilistic decision-making: Mean-field predictions and spiking simulations

Daniel Martí^{a,*}, Gustavo Deco^{a,b}, Paolo Del Giudice^c, Maurizio Mattia^c

^a*Computational Neuroscience Unit, Universitat Pompeu Fabra, Pg. Circumval·lació 8, E-08003 Barcelona, Spain*

^b*Institució Catalana d'Estudis Avançats (ICREA)*

^c*Complex Systems Unit, Department of Technologies and Health, Istituto Superiore di Sanità, V.le Regina Elena 299, 00161 Roma, Italy*

Abstract

In this work we study the basic competitive and cooperative mechanisms of neural activity in the context of a two-alternative free-choice eye-movement task, as a function of the expectation of reward. We use a simplified version of the protocol followed by Platt and Glimcher [Neural correlates of decision variables in parietal cortex, *Nature* 400 (1999) 233–238], in which each choice is associated with independent underlying reward schedules, and explicitly model it using a biophysically realistic network of integrate-and-fire neurons that forms a categorical choice from the expected gain contingencies, via a simple bias mechanism. The model accounts for several experimental findings, such as the gain-modulated firing activity observed by Platt and Glimcher and the matching law.

© 2006 Published by Elsevier B.V.

Keywords: Computational neuroscience; Decision-making; Network model; Lateral intraparietal area

1. Introduction

In recent years the neural signatures that encode behavioral value have been identified, opening the possibility to investigate decision-making at the physiological level. In this work we study the basic competitive and cooperative mechanisms that underlie the neural activity correlated with the decision-making process in primates. In particular, we model the dependence of neural activity on the expectation of reward associated with the eye-movement response performed by the monkey. To achieve this, we explicitly model the processes occurring at the level of AMPA, NMDA and GABA synapses using a cortical recurrent network of integrate-and-fire neurons. Due to the rich phenomenology of the spiking dynamics, a preliminary analysis of the dynamical regimes accessible to the system is done. This analysis consists in exploring the stationary attractors in the relevant parameter space via a mean-field reduction consistent with the underlying synaptic and

spiking dynamics [4]. Once the regimes of operation of the network are characterized and the corresponding parameter ranges revealed, both the non-stationary dynamical behavior, as measured in neuronal recording experiments, and the asymptotic stationary regimes are studied via the full simulation of the spiking network.

2. Behavioral task

In our simulations we have used a simplified version of one of the protocols used by Platt and Glimcher [5]. In the task, while the subject is keeping his gaze aligned to the fixation point, two eccentric stimuli are illuminated. After some time, the extinction of the central stimulus instructs the subject to look at either of the two eccentric stimuli. The expected gain associated to each stimulus is manipulated by delivering different amounts of juice to the monkey. In this sense the estimation of value made by the subject can be controlled externally. The expected gain for each response is then varied across blocks of trials to test whether neural activity in lateral intraparietal (LIP) area is correlated with subjective value. The frequency with which the animal chooses each response is used as a

*Corresponding author.

E-mail addresses: daniel.marti@upf.edu (D. Martí), gustavo.deco@upf.edu (G. Deco), paolo.delgiudice@iss.infn.it (P.D. Giudice), mattia@iss.infn.it (M. Mattia).

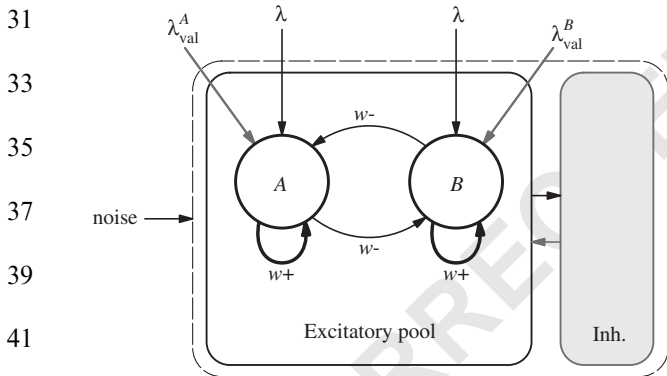
1 behavioral readout of the subjective value of each option.
 2 Platt and Glimcher explicitly proved that subjective value
 3 was represented by the firing activity of LIP neurons. They
 4 also observed the ‘matching law’ in action, giving a linear
 5 dependence of the probability of a given choice on the
 6 reward bias.

7

9 3. Computational model

11 Our network model is composed of a pair of neural
 12 excitatory ‘selective’ populations (or pools, labeled A and
 13 B) with strong recurrent synaptic self-couplings and weak
 14 mutual excitation. In addition, A and B are reciprocally
 15 connected to an inhibitory population and to an ‘un-
 16 selective’ excitatory population; all populations receive
 17 external excitatory synaptic inputs coding for stimuli and
 18 other external influences, including background sponta-
 19 neous activity (see Fig. 1 and [1,6] for the general
 20 theoretical setting). A and B can be ‘selective’ in that they
 21 react to stimuli and can engage in competition due to
 22 shared inhibition, such that even for equal or very similar
 23 inputs to A and B the network can exhibit high A firing
 24 activity with suppressed B activity (which will be taken to
 25 encode ‘decision A’) or the reverse (‘decision B’) [3,7].

27 In absence of stimuli, every cell in the module receives
 28 external input modeled as a Poisson train with rate
 29 $v_{\text{noise}} = v_{\text{out}} N_{\text{ext}} \sim 3 \text{ Hz} \times 800 = 2.4 \text{ kHz}$, where v_{out} is the
 30 average firing rate of any neuron outside the module and



31 Fig. 1. Architecture of the network. The module consists in N neurons
 32 and is divided in two major groups: one inhibitory and one excitatory
 33 population, with N_I and N_E neurons each ($N = N_E + N_I$). Within the
 34 excitatory pool there are two types of populations: two selective pools A
 35 and B, each constituted by fN_E neurons ($f = 0.15$), and one non-selective
 36 population, formed by all excitatory neurons not belonging to a selective
 37 pool ($(1 - 2f)N_E$). The firing rate activity of the two selective pools encode
 38 the decision to make. w_+ are the synaptic weights connecting neurons
 39 within the same selective pool, whereas w_- denotes the connection weight
 40 between neurons in different selective pools and from non-selective to
 41 selective neurons. All other possible connections have weight 1 (baseline
 42 strength). To assure that the overall recurrent excitatory synaptic drive in
 43 the spontaneous state remains constant as w_+ is modified, w_- is set to
 44 $1 - f(w_+ - 1)/(1 - f)$. The signals associated to the stimuli are denoted by
 45 λ and the signal carrying value information is $\lambda_{\text{val}}^{A,B}$ (one per pool). Every
 46 neuron in the network receives a background Poisson spike train of rate
 47 2.4 kHz.

50 N_{ext} is the number of external synapses. The presence of a
 51 stimulus is implemented by an increase of the external
 52 input perceived by every selective neuron. So, during
 53 stimulus presentation a selective neuron, either in A or in
 54 B, receives a Poisson spike train of rate $v_{\text{ext}} = v_{\text{noise}} + \lambda$,
 55 where λ represents the intensity of the stimulus. The
 56 expectation of reward is implemented extrinsically; we
 57 assume that the decision-making process is triggered by an
 58 external signal coming from a module that stores the
 59 representation of value. The value signal is added to the
 60 total background noise perceived by each neural popula-
 61 tion. Even though this is an oversimplified model, it can
 62 shed light on how basic reward-biased decision-making
 63 mechanisms work in a network model.

73 3.1. Mean-field parameter exploration

75 Spiking simulations are too computationally expensive
 76 for an extensive search in the parameter space. Mean-field
 77 approximations allow to compute the attractors to which
 78 the network would converge in the limit of an infinite
 79 number of neurons, and require much less computational
 80 load. The mean-field approximation we used was that
 81 derived by Brunel and Wang [2]. The goal of these
 82 explorations was to find the number of stable network
 83 states (attractors) that coexist for a given set of parameters.
 84 We can distinguish four different stable network states: in
 85 the *spontaneous* state (S) the firing rates of the two
 86 populations are comparable and low ($v_A \simeq v_B \sim 3 \text{ Hz}$); in
 87 the *mixed state* (M) the activity of the two populations is
 88 also comparable, but substantially higher than typical
 89 spontaneous activity. The other two states are the *selective*
 90 states (A and B), in which one of the two populations
 91 shows elevated activity while the other population fires at a
 92 very low (suppressed) rate. In a selective state the ratio of
 93 the high firing rate over the suppressed firing rate,
 94 $v_{\text{high}}/v_{\text{low}}$, is typically higher than 10. If the stable states
 95 A and B represent the two categorical options the network
 96 has to choose from, the mixed state M could be an
 97 interesting dynamic option for describing an ‘undecided’
 98 state, possibly corresponding to unusually long decision
 99 times.

101 The network shows multistable behavior: there may
 102 coexist several stable states given a fixed set of parameters.
 103 The set of stable states that the network can sustain for
 104 some values of the parameters determines the phase or
 105 regime of operation of the system. In this system, the
 106 relevant parameters are the recurrent potentiation weight
 107 w_+ , and the amplitude of the external signal received by
 108 each neuron in a pool: $v_{\text{ext}}^A = v_{\text{noise}} + \lambda + \lambda_{\text{val}}^A$,
 109 $v_{\text{ext}}^B = v_{\text{noise}} + \lambda + \lambda_{\text{val}}^B$. It is convenient to define $\bar{\lambda} \equiv \lambda +$
 110 $(\lambda_{\text{val}}^A + \lambda_{\text{val}}^B)/2$ and $\Delta\lambda \equiv (\lambda_{\text{val}}^A - \lambda_{\text{val}}^B)/2$. Fig. 2 summarizes
 111 the regimes of operation of the network found at each
 112 point in the parameter space.

113 The network should take a binary decision between A
 114 and B once the value information given by both λ_{val}^A and
 115 λ_{val}^B is available, rather than being trapped in a non-

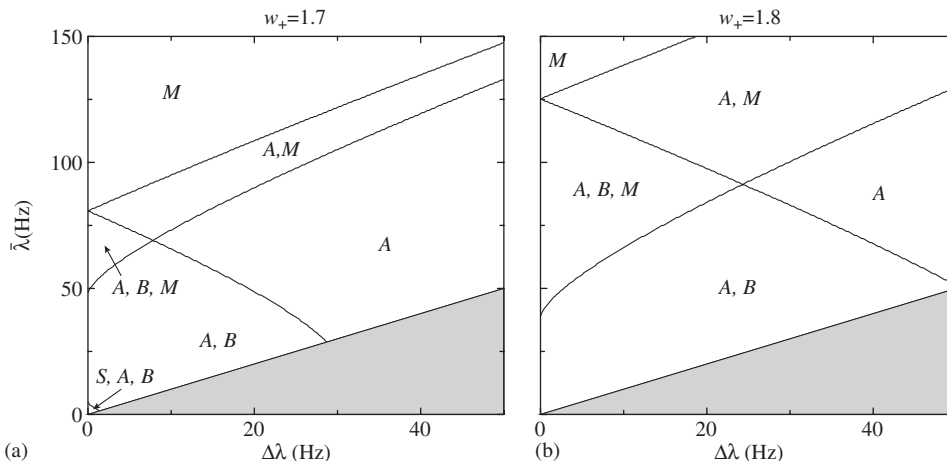


Fig. 2. Mean-field exploration of the network's regimes of operation for $w_+ = 1.7$ and $w_+ = 1.8$. Each region corresponds to a regime of operation, determined by the set of stable states that the network can sustain. Stable states are denoted by their initials: S (spontaneous), M (mixed), A (A-selective), and B (B-selective). The shaded area represents the unphysical region, defined by $\Delta\lambda > \bar{\lambda}$. In the competition region (A,B) there are two stable selective states, and the network must take a binary decision. Note that the diagrams show only the region $\Delta\lambda > 0$; the diagrams for $\Delta\lambda < 0$ are the mirror images about $\Delta\lambda = 0$ of the diagrams shown here, with A and B interchanged.

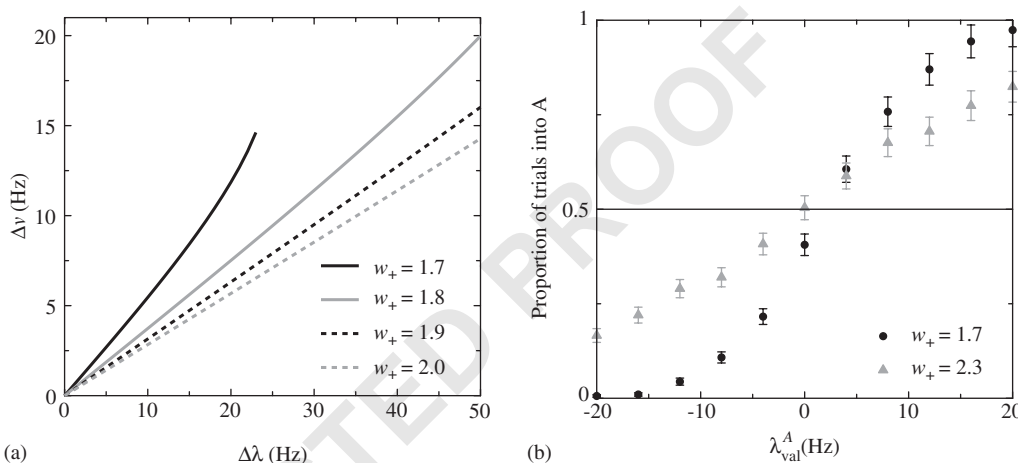


Fig. 3. (a) Difference between the asymptotic firing rates associated with a high and a low expected gain, Δv , with respect to the bias in the value signal, $\Delta\lambda$, for different values of the recurrent potentiation weight w_+ and with $\bar{\lambda} = 0$. The curves are plotted in the $\Delta\lambda$ interval where there is competition. (b) Proportion of trials (out of 500) in which the network chose A as a function of the value signal received at A (λ_{val}^A), keeping the other value signal (λ_{val}^B) fixed at 0. Negative values for λ_{val}^A should be interpreted as the symmetric situation, where only population B receives a value signal, equal to $\lambda_{\text{val}}^B = |\lambda_{\text{val}}^A|$.

selective state, where no decision is made. Since we wanted the network to choose between A and B, we were specifically interested in finding a region of multistability in the parameter space where the only stable states were the two selective ones. This region is called the *competition region*, and provides (in an approximate way) the parameter ranges to use in the spiking simulations. As seen in Fig. 2, the competition region widens as w_+ increases. This widening reflects the fact that the sensitivity to the bias is reduced whenever the self-excitation is increased. Higher values of w_+ require then a stronger bias to destabilize the B state.

3.2. Spiking dynamics

Once the region in parameter space that allows a pure *binary-decision* was found, we used an explicit simulation of the network's spiking dynamics. The number of neurons in the network was $N = 1000$. Simulations followed the same protocol of the behavioral task we modeled. During the first 500 ms λ and λ_{val} were set to 0 Hz to represent the absence of competing stimuli in the first stage of the psychophysical task; after cue onset, and until the rest of the task, all neurons in both A and B pools received, apart from the stimulus-driven λ signal, a pool-specific value signal, λ_{val}^A and λ_{val}^B . We calculated the difference between the asymptotic firing rate of the winner population under high- and low-gain conditions (i.e., favored and unfavored

1 by the bias), or *modulation* (Δv), for different values of the
 3 the fraction of choices into A. Fig. 3 summarizes the results
 5 obtained, and shows a qualitative agreement with the
 7 analysis of Platt and Glimcher. In particular, we note from
 9 Fig. 3a that the amplitude of the modulation Δv depends
 11 linearly on the bias $\Delta \lambda$ over a wide range, parallel to
 13 observations in Platt and Glimcher. It is also seen that the
 15 slope of the straight lines are strongly modulated by w_+ :
 17 increasing the self-excitation w_+ lowers the sensitivity of
 19 the modulation $\Delta \lambda$ to the value bias $\Delta \lambda$. This loss of
 sensitivity to the bias when w_+ increases results from the
 dominance of recurrent synaptic components over the
 external inputs. This effect is also observed in Fig. 3b,
 which shows to what extent the decision outcome is
 sensitive to the reward signal, for different values of the
 recurrent excitation. One could speculate that plasticity
 induced in the w_+ synapses could provide a way for tuning
 the choice performance for a given value information.

21 References

- 23 [1] D.J. Amit, N. Brunel, Model of global spontaneous activity and local
 25 structured activity during delay periods in the cerebral cortex, *Cereb.*
Cortex 7 (3) (1997) 237–252.
 27 [2] N. Brunel, X.J. Wang, Effects of neuromodulation in a cortical
 network model of object working memory dominated by recurrent
 inhibition, *J. Comput. Neurosci.* 11 (1) (2001) 63–85.
 29 [3] G. Deco, E.T. Rolls, Synaptic and spiking dynamics underlying
 reward reversal in the orbitofrontal cortex, *Cereb. Cortex* 15 (1) (2005)
 15–30.
 31 [4] P. Del Giudice, S. Fusi, M. Mattia, Modeling the formation of
 working memory with networks of integrate-and-fire neurons con-
 nected by plastic synapses, *J. Physiol.-Paris* 97 (4–6) (2003) 659–681.
 33 [5] M.L. Platt, P.W. Glimcher, Neural correlates of decision variables in
 parietal cortex, *Nature* 400 (6741) (1999) 233–238.
 35 [6] E.T. Rolls, G. Deco, *Comput. Neurosci. of Vision*, Oxford University
 Press, Oxford, 2003.
 37 [7] X.J. Wang, Probabilistic decision making by slow reverberation in
 cortical circuits, *Neuron* 36 (2002) 955–968.



worked on computational problems in radiotherapy, and the statistical
 analysis of DNA sequences.



proaches.

Paolo Del Giudice graduated in physics in 1985
 from the Rome University “La Sapienza”. Since
 1991 he works at the Italian National Institute of
 Health (Complex systems unit of the Technolo-
 gies and Health Department). He has been
 mostly active in the theory, simulation and
 electronic implementation of neural network
 models, recently focusing in particular of the
 collective stochastic dynamics of spiking neurons
 and neuromorphic multichip systems. He also
 worked on computational problems in radiotherapy, and the statistical
 analysis of DNA sequences.

Maurizio Mattia received his degree in physics
 from the University of Rome “La Sapienza” in
 1997. He is a researcher at the Department of
 Technologies and Health of the Italian National
 Institute of Health (Istituto Superiore di Sanità),
 and works on the understanding and control of
 the behaviour of spiking neuron populations
 treated as stochastic nonlinear systems. He is
 also active on medical physics issues dealt with
 complex system theory and computational ap-



Daniel Martí graduated in Physics (UAB, Barce-
 lona) and Theoretical Particle Physics (M.Sc.,
 UAB-IFAE, Barcelona) in 2003. He is currently a
 Ph.D. student in the Computational Neuro-
 science Group at Universitat Pompeu Fabra,
 Barcelona. His research interests include prob-
 abilistic models of perception and decision, and
 population dynamics of network of spiking
 neurons.



Gustavo Deco is a Research Professor from the
 Institució Catalana de Recerca i Estudis Avanç-
 ats at the Universitat Pompeu Fabra, Barcelona,
 where he is leading the Computational Neuro-
 science Group. He studied physics at the
 Rosario National University (Argentina), and
 he received his Ph.D. degree in Physics in 1987.
 His research interests include models of visual
 attention, decision-making, and neural dynamics.