Flexible Control of Mutual Inhibition: A Neural Model of Two-Interval Discrimination
Science 307, 1121 (2005);
DOI: 10.1126/science.1104171

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 8, 2008):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/307/5712/1121

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/307/5712/1121/DC1

A list of selected additional articles on the Science Web sites related to this article can be found at:
http://www.sciencemag.org/cgi/content/full/307/5712/1121#related-content

This article cites 29 articles, 8 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/307/5712/1121#otherarticles

This article has been cited by 39 article(s) on the ISI Web of Science.

This article has been cited by 19 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/307/5712/1121#otherarticles

This article appears in the following subject collections:
Neuroscience
http://www.sciencemag.org/cgi/collection/neuroscience

Information about obtaining reprints of this article or about obtaining permission to reproduce this article in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl
Flexible Control of Mutual Inhibition: A Neural Model of Two-Interval Discrimination

Christian K. Machens,1 Ranulfo Romo,2 Carlos D. Brody1*

Networks adapt to environmental demands by switching between distinct dynamical behaviors. The activity of frontal-lobe neurons during two-interval discrimination tasks is an example of these adaptable dynamics. Subjects first perceive a stimulus, then hold it in working memory, and finally make a decision by comparing it with a second stimulus. We present a simple mutual-inhibition network model that captures all three task phases within a single framework. The model integrates both working memory and decision making because its dynamical properties are easily controlled without changing its connectivity. Mutual inhibition between nonlinear units is a useful design motif for networks that must display multiple behaviors.

In our daily lives, our minds can flit from thought to thought with remarkable speed and flexibility (1). A simplified task that requires rapid shifts between different mental actions is known as two-interval discrimination (two stimuli separated by a time interval; Fig. 1A). Subjects must first perceive a brief stimulus, called f1, maintain it in working memory for several seconds, and then compare it with a brief second stimulus, called f2, to immediately decide which of the two stimuli was larger. The task requires both working memory and decision making, interfacing between the two in a rapid switch from one to the other.

The biophysical mechanisms underlying the performance of this task remain unknown.

1Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA. 2Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, 04510 México, D.F., México.

*To whom correspondence should be addressed. E-mail: brody@cshl.edu

Figure 1, C and D show the firing rates of two prefrontal cortical (PFC) neurons recorded from previously trained macaque monkeys while they performed a two-interval discrimination task in which f1 and f2 were the frequencies of mechanical vibrations applied to the tip of a finger (12, 16, 22). The dynamics of the activity of these neurons depends strongly on the phase of the task. During the loading of f1 into working memory, there is a rapid flow to an f1-dependent firing rate. During the maintenance of f1 in working memory, there is a long-lasting persistence of f1-dependent firing rates, despite the absence of the stimulus. During the comparison/decision phase, upon presentation of the second stimulus f2, the firing rates quickly segregate into one of two categories, depending on the monkey’s subsequent choice of a “yes” or “no” push-button answer to the question, “Is f1 greater than f2?” Responses similar to these PFC responses are also found in ventral (17) and medial (14) premotor cortices. For brevity, here we will refer collectively to these three areas as “frontal lobe areas.” We highlight two aspects of the frontal lobe data. First, signals are encoded in complementary sets of roughly equal numbers of neurons (12, 14, 17). One set is comprised of “plus” neurons, defined as neurons with a delay-period firing rate that is a monotonically increasing function of f1 (Fig. 1C). Plus neurons typically fire the most for “yes” decisions after presentation of f2. The complementary set are “minus” neurons, defined as those which have delay period firing rates that are monotonically decreasing functions of f1, and fire the most for “no” decisions (Fig. 1D). Because higher f2 stimuli are more likely to lead to “no” decisions; plus neurons are excited by high f1 stimuli but inhibited by high f2 stimuli. The converse occurs for minus neurons.
Each of the two sets of neurons carries all the information necessary for the task; the existence of two apparently redundant sets is so far unexplained.

Second, the same neurons that show graded delay-period activity, which represents working memory of \( f_1 \), also show categorical activity after stimulus \( f_2 \), which represents the monkey’s decision (14, 17, 24). This finding contrasts with current mathematical psychology models of two-interval discrimination, which have implicitly assumed that working memory and decision making are processes represented by separate variables (25–28). Instead, we propose an algorithm (Fig. 2) in which both memory and decision outcome are represented by the value of a single state variable (horizontal axis). The dynamical modes of the system are described by a hypothetical energy function \( L \) (vertical axis), the shape of which does not depend on the value of the state variable. The state variable always evolves so as to reduce \( L \). During the loading phase, the external stimulus creates a single minimum in \( L \) (a single stable point) at a position determined by the value of \( f_1 \). This forces the state into an \( f_1 \)-dependent position. During the memory maintenance phase, there is no longer an external stimulus that determines the shape of \( L \). The memory of \( f_1 \) is represented by the state’s position, and for this to remain steady, the \( L \) function must be approximately flat [a line attractor configuration (2, 29, 30)]. During the comparison/decision phase, we map stimulus \( f_2 \) onto the same horizontal axis as \( f_1 \). All state positions to the left of \( f_2 \) now represent memories of \( f_1 \) values that are less than \( f_2 \). States to the right of \( f_2 \) represent memories of \( f_1 \) that are greater than \( f_2 \). If a peak in \( L \) (an unstable point) is created at the position given by \( f_2 \), the state will evolve in one of two opposite directions, depending on the yes or no answer to \( f_1 > f_2 \)? If the horizontal axis stands for a firing rate that grows from left to right, the plots mimic the activity of plus neurons through all phases of the task. If the firing rate grows from right to left, the plots mimic minus neurons.

We propose that the frontal-lobe areas instantiate this dynamical algorithm. The task-relevant sensory inputs to these frontal areas arise from the secondary somatosensory cortex (S2) (Fig. 1, E and F) (31). During the first stimulus, responses in the S2 are similar to those in the PFC: there are both plus (Fig. 1E) and minus (Fig. 1F) neurons (13, 15). Unlike the PFC, however, during most of the delay period, S2 neuron firing rates are low and not \( f_1 \)-dependent. Also in contrast with the PFC, immediately after presentation of \( f_2 \), neurons in the S2 respond to \( f_2 \) with the same plus-or minus-type firing-rate dependence with which they responded to \( f_1 \) (32).

Clues about the underlying frontal-lobe neural architecture come from the analysis of firing rate covariations between pairs of PFC neurons. These tend to be positive if both neurons are plus, or if both are minus, but negative when one is plus and one is minus (33), leading us to consider the architecture sketched in Fig. 3A. In a simplified version of this circuit (Fig. 3B), each node represents a population of neurons, and each node’s output variable is the average activity of the population. Figure 3C shows a node’s i/o function, which is defined as its output, expressed as a resulting postsynaptic conductance, as a function of excitatory input (34). The overall shape obtained, with a threshold below which output is negligible and with saturation at high outputs, is characteristic of many neuron models.

To study graphically the dynamics of the circuit in Fig. 3B, we show in black in Fig. 3D the output of the plus node as a function of the inhibitory input from the minus node. An additional excitatory input, \( E \), is held constant here. The minus node’s i/o function can be plotted by exchanging the horizontal and vertical axes to form the brown axes and curve in Fig. 3D. This phase-plane plot now describes the complete dynamics of the system, because we can follow the input to output activity of each node as it reverberates around the circuit loop. Points where the two i/o curves intersect are known as fixed points, or steady states. These may be stable (like the minimum in \( L \) during loading as seen in Fig. 2) or unstable (like the maximum in \( L \) during comparison/decision).

During the first stimulus [loading phase (Fig. 3E)], inputs from the S2 area are active. Together with the external input \( E \), they can shift the i/o functions along their input axes. If both nodes receive the same S2 input, then by symmetry, the crossing point of the two i/o curves must lie along the 45° diagonal of the phase-plane plots. However, we propose that S2 plus neurons project to frontal plus neurons, and S2 minus neurons project to frontal minus neurons. As a result, the position of the single stable point is determined by the value of \( f_1 \), thus instantiating the loading mode of Fig. 2. Horizontal position in Fig. 2 corre-
sponds here to the angle of \( \theta \) (defined in top phase plane, Fig. 3E).

During the delay period (maintenance phase (Fig. 3F)), the inputs from S2 are silent. The value of \( E \) and the neuronal-model parameters can be chosen such that the two i/o curves largely overlap. This creates a quasi-continuous line of stable points (a line attractor) and thus instantiates the maintenance mode of Fig. 2 (35).

During the comparison/decision phase (Fig. 3G), we propose that an external control signal, which indicates that the current stimulus should be treated as \( f_2 \), not \( f_1 \), reduces the otherwise constant excitatory input \( E \). For low enough \( E \), the system has two stable fixed points on either side of a single unstable fixed point, as required for this phase in Fig. 2. In the data, frontal neurons switch the sign of their stimulus dependence between \( f_1 \) and \( f_2 \) (25), but S2 sensory neurons do not. The net functional connection between S2 and frontal neurons must therefore change sign. To match this in the model, we used the circuit shown in Fig. 3H to switch the net S2-to-frontal plus neuron connections between loading (Fig. 3E, top) and comparison/decision (Fig. 3G, top). A similar circuit was used for frontal minus neuron connections (36). As a result of this input sign switch, increasing \( f_2 \) moves the unstable fixed point from lower to higher \( \theta \) (Fig. 3G), which matches the \( f_1 \)-dependence of the stable fixed point during loading (Fig. 3E). This completes instantiation of the comparison/decision mode of Fig. 2.

The continuous-variable nodes of the model of Fig. 3 were each replaced by 250 noisy, leaky, integrate-and-fire neurons in order to produce a spiking neuron model with almost identical behavior (Fig. S4). Figure 4A shows firing rates for one spiking neuron from the plus node and one neuron from the minus node, qualitatively capturing key aspects of the data. The sign of correlations between pairs of neurons in the model also matches the pattern found in the experimental data (Fig. 4B).

The one-dimensional algorithm of Fig. 2 produces testable predictions separate from the neural instantiation proposed here. If the decision is reported when the state reaches one of the two final stable points in Fig. 2, and if the distance to the final stable point is a dominant factor in the time required to reach it, then it follows that on trials where \( f_1 \) equals \( f_2 \) (50% “yes” responses, 50% “no”) and \( f_1 \) and \( f_2 \) are near the high end of the range of stimuli, “yes” decisions will be reached faster than “no” decisions. The converse is true at the low end of the range (Fig. 4D). At the neural instantiation level, both the fine-tuned (Fig. 3) and robust (34, 35) instantiations rest on populations of mutually inhibitory neurons. If the strength of inhibitory connections were increased [e.g., by use of benzodiazepines (37)], then the memory maintenance mode would be perturbed in the direction of the comparison/decision mode, with a single unstable point. Because of the circuit’s symmetry, the effect would be similar to an \( f_2 \) stimulus in the middle of the range (\( \approx 22 \) Hz), even though no stimulus would be present (Fig. 4C). The psychophysical correlate would be a tendency to categorize \( f_1 \) into high or low rather than remembering it accurately.

Unlike previous line attractor models (3, 5, 6), which relied mostly on mutual exci-

![Fig. 3. Mutual inhibition model implements the dynamical modes of Fig. 2. (A and B) Circuit diagrams. Each node represents a population of neurons. (C) Input/output of a node as a function of excitatory input (nS, nanoSiemens). (D) Phase-plane plot of both i/o functions allows tracing out the dynamics of the mutual inhibition circuit. (Black arrows indicate stepping discretely between nodes, for illustration, and gray arrows indicate continuous time dynamics, as used in actual simulations.) (E) Loading. The top panel shows inputs to the circuit. The lower three panels show the effect of three different values (10, 22, and 34 Hz) of the first stimulus \( f_1 \). Colored lines show dynamical trajectories, starting from \( (0,0) \) and ending at filled circles. (F) Maintenance. Removing inputs from S2 results in a quasi-continuum of stable points. Dynamical trajectories start from correspondingly colored endpoints of trajectories in (E). (G) Comparison/decision. The lower panels show i/o functions and dynamics for \( f_2 = 18 \) (top) and \( f_2 = 26 \) (bottom). The final state depends on the answer to \( f_1 > f_2 \)? Trajectories start from the endpoints of (F). For all plots in (E), (F), and (G), angle \( \theta \) corresponds to horizontal position in Fig. 2. (H) Input switching circuitry. At high \( E \), S2 plus signals pass through, but at low \( E \), S2 minus signals pass through.](www.sciencemag.org)
tation, we implemented both working-memory and decision-making dynamics through mutual inhibition. This facilitated the use of a single simple circuit for both. The use of neuron models with i/o functions that are nonlinear above threshold, instead of linear or threshold-linear models (38–40), further allowed straightforward control of the circuit's dynamical mode through an external excitation signal.

Our daily mental lives have an enormous variety of highly flexible dynamics. What is the neural basis of this flexibility? Do frontal lobes contain many separate modules of neurons, each capable of a particular type of computation and its attendant dynamics? Or, as the data of Fig. 1, C and D have inspired us to propose here, can single modules of frontal lobe neurons rapidly reconfigure their dynamical properties, switching between different behaviors as the cognitive flow requires? If single modules are indeed flexible, what is the range of dynamics and computations that they can display? We have only begun to address these questions here. But we believe they are fundamental, and lie at the heart of the nature of the neural architecture underlying cognition.

References and Notes

23. PFC neurons change the net sign of their stimulus dependence between stimuli f1 and f2. That is, the higher the stimulus value applied during comparison/decision, the lower the PFC neuron's firing rate and the higher the PFC minus neuron's firing rate because higher f2s are more likely to lead to “no” decisions (fig. 5).
32. During the late part of the second stimulus, the firing of many S2 neurons becomes correlated with the monkey's decision [15], but this occurs after decision-correlated activity arises in the PFC, and thus cannot be causal to it. For simplicity, in the model we treat only the initial creation of decision-correlated activity in the PFC, before it has propagated back to area S2.
33. During the delay period and across trials using identical f1 stimuli, we calculated the noise correlation r between pairs of neurons [41]. We found covariation between pairs of plus neurons r = 0.13 ± 0.04 (56), n = 32 – pairs, and minus neurons r = -0.08 ± 0.03, n = 17. But pairs where one neuron is plus and the other minus were anticorrelated r = -0.04 ± 0.02, n = 58 [24].
34. Materials, methods, and documented computer code are available as supporting material on Science Online.
35. Good overlap in Fig. 3F requires high precision in the design of the i/o function shapes. This is a generic difficulty that all line attractors face [30, 42]. We address this issue in the supporting online text by adapting the solution proposed by Koulakov et al. [6] to create a more robust model. Despite its greatly increased dimensionality, the more robust model can still be approximately described in terms of two variables (fig. S6), similar to the diagrams of Fig. 3.
36. The circuit of Fig. 3H is only one possible way to switch input signs. See [43] for another possibility.
46. C.K.M. and C.D.B. jointly developed and implemented the model and wrote the paper. R.R. conceived and carried out the experiments. This work was supported in part by a Swartz Foundation Fellowship to C.K.M.; an international Research Scholars Award from the Howard Hughes Medical Institute and awards from Dirección General de Asuntos del Personal Academico–Universidad Autonoma Nacional de Mexico and the Millennium Science Initiative–Consejo Nacional de Ciencia y Tecnologia to R.R.; and by a Sloan Foundation Research Fellowship, an award from the Redwood Neuroscience Institute, and NIH grant 1R01MH067991-01 to C.D.B. We thank J. D. Cohen, Z. F. Mainen, S. S. H. Wang, and A. M. Zador for comments on the manuscript, and J. J. Hopfield for discussion.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5712/1121/DC1
Materials and Methods
Figs. S1 to S9
References
Model S1

17 August 2004; accepted 6 January 2005
10.1126/science.1104171

Downloaded from www.sciencemag.org on April 8, 2008