Federico J. Carnevale

Neural Dynamics of Perceptual Detection under Temporal Uncertainty

To my father,

Abstract

Among the large number of functions that compose our mental life, perception is arguably the most fundamental one. Perception is the cognitive process by which external sensory signals are transformed into meaningful information that represents our environment and guide decisions and behavior. What is the neural basis of this transformation? Several key issues limit our understanding of the neurobiology of perception and perceptual decision-making. First, the neural codes used by the brain to represent sensory information are still unclear. Second, perceptual decisions presumably arise from the coordinated activity of populations of neurons. However, the analytical tools best suited to study decision signals in neuronal populations remain unknown. Third, perception is not a passive process. On the contrary, external stimuli and internal brain states dynamically interact to give rise to percepts. In this thesis, I address these questions using computational simulations and neural data recorded while monkeys perform a vibrotactile detection task. Three fundamental issues are examined: (1) the dynamics of correlated variability, (2) the decoding of decisions from neural population's activity and (3) the neural mechanisms underlying the use of temporal expectations. I study the dynamics of choice-conditioned noise correlations and show that they reveal an internal component of the decision-making process. By developing novel statistical measures, I quantify how predictive is the activity of populations of cortical neurons about the subject's decision. As a result, I find that a specific subset of premotor cortex neurons unequivocally predict the animal's decision report. The vibrotactile detection task studied in this work requires subjects to make decisions under

temporal uncertainty. I find that subjects benefit from temporal expectations by modulating their response criterion over the course of a trial. I show that this modulation is represented by the population dynamics of premotor cortex neurons. A trained recurrent neural network reproduces the experimental findings and reveals the dynamical mechanism implementing a flexible response criterion. Knowledge about the probability of stimulation over time, acquired during training, is intrinsically encoded in the neural population activity, allowing a dynamic control of the response criterion to improve performance.

Resumen

Entre el gran número funciones cognitivas que componen nuestra vida mental, la percepción es, quizá, la más fundamental. La percepción es el proceso mediante el cual el cerebro interpreta, organiza y da sentido a la gran cantidad de señales sensoriales que recibe del mundo exterior. De esta forma, la información sensorial es transformada en una representación relevante de nuestro entorno, útil para guíar nuestro comportamiento. ¿Cuál es el correlato neuronal de esta transformación? Hay varias cuestiones clave que limitan nuestro entendimiento de la neurobiología de la percepción y de las decisiones perceptuales. En primer lugar, el código neuronal que el cerebro utiliza para representar información sensorial no es del todo claro. En segundo lugar, las decisiones presumiblemente emergen de la actividad conjunta de un gran número de neuronas. Sin embargo, las herramientas analíticas más adecuadas para estudiar estas señales poblacionales todavía no son enteramente conocidas. En tercer lugar, la percepción no es un proceso pasivo. Por el contrario, los estímulos externos y los estados internos del cerebro interactúan dinámicamente para construir nuestra experiencia subjetiva. En esta tesis, abordo estos asuntos utilizando simulaciones computacionales y analizando registros neuronales obtenidos en monos mientras realizan una tarea de detección vibrotáctil. Tres cuestiones fundamentales son examinadas: (1) la dinámica de la variabilidad neuronal correlacionada, (2) la decodificación de señales de decisión a partir de la actividad de poblaciones de neuronas y (3) los mecanismos neuronales que subyacen a la incorporación de expectativas temporales en el proceso de decisión. Estudiando la dinámica de las correlaciones del ruido, muestro que éstan

revelan una componente interna del proceso de decisión. Mediante el desarrollo de nuevas medidas estadísticas, cuantifico el poder predictivo de la actividad de conjuntos de neuronas acerca del las decisiones del sujeto. Como resultado, encuentro que la decisión del animal puede predecirse inequívocamente a partir de la actividad de poblaciones específicas de neuronas de la corteza premotora. La tarea de detección estudiada en esta tesis require que los animales tomen decisiones en un contexto de incertidumbre temporal. En esta tesis muestro que los sujetos construyen y utilizan expectativas temporales para aumentar su rendimiento mediante la modulación de su criterio de respuesta a través del tiempo. Además, encuentro que la actividad de las neuronas de la corteza premotora es consistente con un mecanismo neuronal específico para implementar esta modulación. Finalmente, derivo un modelo de red recurrente que reproduce los resultados experimentales y permite estudiar la estructura dinámica subyacente. El conocimiento previo acerca de la probabilidad de estimulación como función del tiempo, adquirido durante el entrenamiento, puede ser intrínsecamente codificado por una población de neuronas, permitiendo el control dinámico del criterio de durante el proceso de decisión.

Contents

Chapter 1

Introduction

"Men ought to know that from nothing else but from the brain come joys, delights, laughter and sports, and sorrows, griefs, despondency, and lamentations. And by this, in an special manner, we acquire wisdom and knowledge, and see and hear, and know what are foul and what are fair, what are bad and what are good, what are sweet and what unsavory."
— Hippocrates, *On the sacred disease* (400BC)

1.1 Overview

Understanding the brain is at the frontier of modern science. How does this ∼1.5Kg mass of jelly tissue control every single thing we ever think, remember, perceive, feel or do? To realize the amazing capabilities of the human brain there is no need to go very far. As Steven Pinker puts it, "we should direct out awe not at Shakespeare or Mozart or Einstein or Kareem Abdul-Jabbar but at a four-year old carrying out a request to put a toy on a shelf". From an engineering perspective, the problems that the brain solves every day are astonishing. Even more importantly, understanding the brain is essential to comprehend more subtle concepts which are fundamental for our mental life like sensations, self-awareness, free will or consciousness. The ultimate goal of cognitive neuroscience is to provide

satisfactory scientific answers to how neural circuits achieve these capabilities and shape our everyday experience of the world.

1.2 Perception and perceptual decision-making

Among the large number of functions that compose our mental life, perception is perhaps the most fundamental one. Perception is the cognitive process by which we interpret the massive amount of external stimuli that our brain receives from the world. It deals with transforming the signals captured by our peripheral sensors to meaningful representations of the environment (Romo and de Lafuente [2013]). Noisy sensory information, forming an incomplete picture of an uncertain and ambiguous world, is encoded as levels of voltage in sensory receptors. From this information, the brain needs to produce accurate unified percepts to serve as the basis for making decisions and control behavior.

Closely linked to perception is perceptual decision-making: how does the brain use sensory stimuli to guide behavior? Perceptual decision-making studies the cognitive link between sensation and action (Romo and Salinas [2001]). While in many cases the purely sensory and motor components are well understood, intermediate decision-making steps remain unclear. This is mainly a matter of complexity: when deciding in an uncertainty environment, the brain not only has to weight the incoming sensory evidence but also integrate it with many other internal factor as values, expectations, priors, rules, etc (Gold and Shadlen [2007]). Moreover, these components cannot be easily separated or experimentally controlled. Thus, the mechanics by which these factors are represented in the brain and how they are integrated in the decision process are still unknown.

Why to study perceptual decision-making? The importance of understanding the neurobiological basis of perceptual decision-making lies in its capacity to find the correspondences between the mental and neural worlds in terms of neural representations and computations. Perceptual decision-making includes many fundamental unsolved problems: the nature of neural representation (encoding and decoding of information), the origin and impact of neural variability, the

role of behavioral variability, the incorporation of non-sensory influences, etc. Moreover, perceptual decision-making studies many issues that are closely related to the problems of understanding consciousness, at least the *easy* ones (Chalmers [1995]): the ability to discriminate, categorize, and react to environmental stimuli, the integration of information, the reportability of mental states, the ability to access internal states, the focus of attention, the deliberate control of behavior. Thus, perceptual decision-making represents a promising arena in which to study the general principles by which neural circuits support cognition (Shadlen and Kiani [2013]).

1.3 Neural variability and noise correlations

Neural activity is noisy: the response of a neuron during multiple repetitions of a task under - up to the experimenter's best control - identical conditions, is variable (Shadlen and Newsome [1998]; Softky and Koch [1993]). The origin of this variability is unclear and may actually lie in a superposition of different - controllable and uncontrollable- phenomena: from unreliable synaptic transmission to variability in the subject's attention or motivation (Renart and Machens [2014]). Isolating these sources is difficult: in fact, the definition itself of what is considered variability is relative, since to define what is 'noise' it is necessary to assume what is the 'signal' (Masquelier [2013]).

One important feature of neural noise is that it is frequently correlated across pairs of neurons. Studying these correlations, referred to as 'noise correlations', has proved to be useful to understand the function of neural circuits. In this context, the importance of noise correlations is twofold. On the one hand, the impact of neural variability on the information capacity critically depends on its correlation structure (Abbott and Dayan [1999]; Sompolinsky et al. [2001] for review see Averbeck et al. [2006]). In the simplest case, uncorrelated variability can be reduced by averaging across neurons, while if the variability is shared across neurons, it cannot be 'averaged away' (Britten et al. [1992]; Zohary et al. [1994]). On the other hand, noise correlations can be useful as a statistical measure by

itself, for example, to detect changes in internal variables or top-down signals that co-modulate the activity of pairs of neurons.

1.4 Decoding perceptual decisions from neural populations

Neurophysiological recordings performed while animals make perceptual decisions, reveal decision-related signals evolving across sensory and frontoparietal cortices (Gold and Shadlen [2001]; Romo and de Lafuente [2013]; Romo and Salinas [2003]). By simultaneously recording neural activity together with the animal's behavior, these experiments allow to study the involvement of neurons of diferent brain areas in the decision-making process.

A frequently used strategy to study the participation of a neuron in the formation of a decision is to quantify the covariation between the neuron's activity and the subject's choice - that is, the probability with which the subject's decision could be predicted from the activity of the neuron. Although this approach has important limitations (Nienborg and Cumming [2009]; Nienborg et al. [2012]), it nevertheless provides considerable information about the properties of neurons in different cortical areas (Hernández et al. [2010]).

Decisions, however, presumably arise from the interaction of multiple neurons or pools of neurons. Therefore, in order to understand how behavioral choices emerge from these circuits, the relevant measurements must come from population variables. In particular, studying decision signals in neural populations requires defining new statistical tools that take into account the correlated activity of multiple neurons.

1.5 Internal states and temporal expectations

Perceptual-guided behavior results from the combination of the processing of external sensory signals with internal states (Gilbert and Sigman [2007]). We

do not simply respond to external stimuli but we have memory and learning capabilities. Thus, we can benefit from past experience and combine present and past information to better make sense of our surroundings.

The relevance of internal states is evident in the attention literature: regardless of the type of attention-orienting task, the common scenario is that the subject's behavior depends not only on the properties of the stimulus but also in its inner attentional state. Attention can modulate neural circuits to take advantage of predictive information to enhance performance. This was traditionally demonstrated by Posner [1980] in the spatial domain: a hint indicating where a stimulus is more likely to appear lead to a significant decrease in its detection time.

The benefits of using predictive information is not restricted to the spatial domain but it can also refer to time (Nobre [2001]; Nobre et al. [2007]). Temporal expectation is the ability to anticipate or predict the timing of future events in order to optimize the interaction with the sensory world. There is increasing evidence that subjects build temporal expectations and use them to improve performance in decision-making tasks. Temporal expectations are built internally and must be combined with the incoming external signals to guide behavior.

1.6 Signal Detection Theory

To address the problems of perception and perceptual decision-making, it is useful to rely onto the mathematical formalism of Signal Detection Theory (SDT, Green et al. [1966]). SDT is a principled model of how to make decisions under uncertain conditions. It was developed to improve radar detectors during World War II, as the optimal way to detect electrical signals corrupted by interference. Now, it provides a mathematical framework to study perceptual decision-making.

The basics of SDT consists in an agent that must make a binary decision about an unknown state of the world. To do so, it relies on some form of evidence that is indicative of this underlying external and unaccessible state. However, the evidence is noisy and the same piece of evidence could, in principle, be observed

in any of the states of the world, making the decision process difficult.

In the context of the behavioral sciences, the canonical example is a perceptual detection task. An animal is instructed to decide about the presence or absence of a stimulus (the two states of the world corresponding to stimulus absent and stimulus present condition). The evidence comes through its sensory system, for example, as the activity of a population of sensory receptors. This evidence is noisy, either because of intrinsic noise in the neural system or due to external variability in the physical phenomenon. To account for this unreliability, the evidence is considered as a random variable and is characterized by two conditional probability distributions: (1) the probability of the evidence given that the stimulus was present and (2) the probability of the evidence given that the stimulus was absent. The statistical parameters of these distributions (means, variances, etc) depend on the state of the world. For example, it might be that, in average, the neural population fires more when the stimulus is present than when it is not. Thus, the distribution of firing rates given that the stimulus is present will be displaced toward higher values. Crucially, these two distributions might overlap, which makes the inverse problem ill-posed.

Perceptual decisions might be influenced by many factors: the amount of noise in the stimulus, the precision of the sensory system, the a priori probability of the two states of the world, the reward/costs related to each choice, etc. One of the key properties of SDT as a theory resides on its ability to separate these factors in two groups by two independent measures: *sensitivity* and *criterion*.

Sensitivity, or discriminability, refers to the accuracy of the representation and its capability to distinguish the stimulus-present and stimulus-absent distributions. Thus, it essentially depends on the amount of overlap between the two distributions and, therefore, on the distance between their means, relative to their widths. A more sensitive system will have less overlap (either because the means are more separated or because their widths are smaller). Therefore sensitivity is increased by a larger signal or lower amount of noise.

On the other hand, the *criterion* refers to the own judgment of the subject and it is related to probability of making each choice, given the evidence. A

common example is when the cost associated with each of the possible choices is very different. In that case, the subject will plausibly be very reluctant to choose the costly alternative. In general, the decision-maker might be more inclined to one or the other alternative depending on cost-benefits trade-offs, the frequency of the signal, individual preferences, etc. It is important to note that (1) the criterion is independent of the sensitivity and (2) there is no criterion that can achieve zero mistakes if the previously mentioned distributions overlap.

The distinction between sensitivity and criterion is useful because both could be, in principle, flexibly modulated by top-down signals or internal states. Sensitivity might be altered by changes in the neural representation of sensory information that increase the signal-to-noise ratio. Reducing noise correlations in a neural code based on a population of similar neurons would be a possibility. The response criterion could also be modulated in a flexible way, without changing the system's sensitivity. In fact, in this thesis I demonstrate that this is the case when monkeys perform a perceptual detection task with variability in the timing of the sensory evidence.

1.7 The vibrotactile detection task

To understand the neural basis of perception and behavior it is necessary to combine two experimental steps: (1) a perceptual task that provides a rigorous framework in which to study perceptual events and (2) the recordings of neural signals. This combined program of psychophysical and neurophysiological experiments has been pioneered by Vernon Mountcastle in the 60's (Mountcastle et al. [1967, 1969]). By measuring the firing activity of cortical neurons to somatosensory stimuli, they compared the sensibility of these neural responses with the subject's ability to detect or discriminate the same stimuli. Following a similar approach, the laboratory of Prof. Ranulfo Romo at the Universidad Nacional Autónoma de México has been studying the neural correlates of somatosensory perception and decision-making for the last two decades. The work presented in this thesis is based on experimental data recorded in his lab.

In this thesis I focus on a particular behavioral paradigm: the vibrotactile detection task (de Lafuente and Romo [2005, 2006]). In this paradigm, monkeys receive a tactile vibration delivered by a mechanical probe to one of their fingertips. The vibration amplitude varies from trial to trial and stimulus-present trials are randomly interleaved with an equal number of stimulus-absent ones. Monkeys are trained to report whether the vibration is present or not and are rewarded for doing the task correctly. Based on the monkey's report, trials are classified into four types: hits and misses in the stimulus-present condition, and correct rejections and false alarms in the stimulus-absent condition. The animal's behavioral responses are monitored along with its brain's activity which represents an ideal setup to study the neuronal mechanisms underlying perception and perceptual decision-making.

The vibrotactile detection task is specially interesting regarding the study of temporal expectations because of its temporal structure. Trials begin with the stimulator indenting the skin of one fingertip, indicating the monkey that a new trial is about to start. The stimulus to be detected arrives (on half of the trials) after a variable prestimulus period (1.5-3.5 s). Following stimulation (or absence of) monkeys have to wait for a 3 s delay period until a cue indicates to report their decision. Because of this temporal structure, no stimulus arrives before 1.5 s or after 3.5 s. The temporal interval of possible stimulation (between 1.5 s and 3.5 s) is not explicitly cued to the animals. However, monkeys can presumably infer, after many training sessions, this particular structure and use this information to enhance performance in the detection task.

1.7.1 Previous findings

Victor de Lafuente and Ranulfo Romo designed the vibrotactile detection task and performed a set of fundamental experiments exploring the neural basis of somatosensory perception and detection across cortex (de Lafuente and Romo [2005, 2006]). Their findings represent the starting point for the work described in this thesis.

They started by recording the activity of primary somatosensory cortex (S1)

neurons. The activity of these neurons varied continuously as a function of the amplitude of the vibration, suggesting that they are involved in the representation of this physical property of the external stimuli. Neurometric functions obtained on the basis of the activity of S1 neurons matched the animals' psychometric curves. Furthermore, the neural activity did not covaried with the animal's perceptual judgment: no significant differences were found between hits and misses trials.

Next, they recorded the activity of premotor cortex (PMc) neurons while animals performed the task. They found that PMc neurons did not show a monotonic response to the stimulus amplitude. In contrast, their firing rate responded in a all-or-none manner. Moreover, their activity strongly covaried with the animal's behavioral responses (stimulus present or absent), suggesting that PMc neurons are more related to the perceptual judgment that to the physical properties of the stimulus. Because the animal communicated its decisions by a motor act, an alternative interpretation was that premotor activity is representing different motor plans. To test this, they performed a set of control experiments in which they exchanged the push buttons and found that reversing the direction of the movements did not change the activity of PMc neurons. Finally, they tested the causal relationship between PMc activity and the subject's choice. By electrical micro-stimulating PMc neurons they demonstrated that the probability of yes responses significantly increased during micro-stimulation trials.

To sum up, in this set of experiments de Lafuente and Romo demonstrated the fundamental and different roles of sensory and frontal lobe neurons during perceptual detection. While S1 seems to keep a neural representation of the sensory stimulus, PMc seems to be fundamental to the formation of perceptual judgments.

1.8 Aim and overview of this work

The objective of this thesis is to investigate the neural dynamics underlying perceptual detection under temporal uncertainty. It focuses on the detection of a

somatosensory stimuli as behavioral paradigm and explores the dynamics of correlated variability, the decoding of a perceptual decision from population activity and the mechanisms behind the use of temporal expectations in the decision process.

Following this introduction, Chapter 2 describes the dynamics of firing activity and correlated variability in premotor cortex neurons. It shows that these quantities reflect an internal component of the decision-making process that is related with the temporal structure of the task. In Chapter 3, the problem of decoding decisions from neural population's activity is addressed. It describes novel mathematical tools to quantify how predictive of the subject's choice is the activity populations of PMc neurons. Chapter 4 deals with how external stimuli and internal states interact during the decision process. I addresses how prior information about the task's temporal structure can be exploited during the decision-making process. Using sophisticated data analysis and modeling, a specific neural mechanism to benefit from temporal expectations is found. Finally, the last chapter describes the main conclusions of this work.

Part of the work presented in this thesis has been published or presented in the following articles and conferences:

- Carnevale F, Barak O, de Lafuente V, Romo R and Parga N
  Dynamic control of response criterion during perceptual detection under temporal uncertainty. *in Preparation*

- Carnevale F, Barak O, de Lafuente V, Romo R and Parga N
  On the neural dynamics of perceptual decision-making under temporal uncertainty. Cosyne: Computational and Systems Neuroscience Conference, 2014

- Carnevale F, de Lafuente V, Romo R and Parga N
  An Optimal Decision Population Code that Accounts for Correlated Variability Unambiguously Predicts a Subjects Choice. Neuron 80 (6), 1532-1543, 2013

- Carnevale F, de Lafuente V, Romo R and Parga N

The role of neural correlations in a vibrotactile detection task. Computational Neuroscience Meeting: CNS 2013, Paris, France.

- Parga N, Carnevale F, de Lafuente V and Romo R
  On the role of neural correlations in decision-making tasks. Cosyne: Computational and Systems Neuroscience Conference, 2013

- Carnevale F, de Lafuente V, Romo R and Parga N
  Internal signal correlates neural populations and biases perceptual decision reports. Proceedings of the National Academy of Sciences USA 109: 18938-18943, 2012

- Carnevale F, de Lafuente V, Romo R and Parga N
  Uncertainty in stimulus amplitude and application time reveals purely internal neural processing in perceptual decision-making tasks. Society for Neuroscience 2012, New Orleans, USA.

Chapter 2

Dynamics of correlated variability in PMc

"Indeed there are now no logical (and I believe no insurmountable technical) barriers to the direct study of the entire chain of neural events that lead from the initial central representation of sensory stimuli, through the many sequential and parallel transformations of those neural images, to the detection and discrimination processes themselves, and to the formation of general commands for behavioral responses and detailed instructions for their motor execution."
— Vernon Mounscastle, *Handbook of Physiology: The Nervous System* (1984)

2.1 Introduction

Animals often make perceptual decisions under uncertain conditions (Cook and Maunsell [2002]; de Lafuente and Romo [2005, 2006]; Gold and Shadlen [2000]; Hanes and Schall [1996]; Hernández et al. [2002, 2010]; Kim and Shadlen [1999]; Romo et al. [1999, 2002, 2004]; Salinas et al. [2000]; Shadlen and Newsome [2001]; Shadlen et al. [1996]). The arrival of a behaviorally relevant sensory stimulus is usually unknown and its presence is often ambiguous because it can be weak and appear in a noisy background. What are the neural mechanisms underlying

the decision-making process in this situation? Neurophysiological experiments often use clear cues indicating when to start gathering the sensory evidence on which decisions are based (Hernández et al. [2010]; Shadlen et al. [1996]). In these paradigms, experimental data can be explained by feedforward accumulation models (Gold and Shadlen [2000]; Hanes and Schall [1996]; Shadlen and Newsome [2001]; Shadlen et al. [1996]). However, when the time of stimulus onset is variable, neural integration of sensory signals is problematic because it would start either too soon, in which case noise will dominate the process, or too late, losing part of the signal. There is evidence that the brain uses internal signals to guide detection of sensory stimuli (Nienborg and Cumming [2009]). These signals are related to task contingencies that prefrontal cortical networks acquire during training (Miller [2000]) and combine with the stimulus to produce the behavioral response following a process different from simple integration of the sensory evidence (Hernández et al. [2010]).

To investigate the neuronal mechanisms coping with uncertainty about stimulus onset and the role of internal signals in sensory perception, we recorded the simultaneous activity of pairs of premotor cortex neurons, while trained monkeys performed a vibrotactile detection task (de Lafuente and Romo [2005, 2006]). In this task, the stimulus was often absent or weak, and the time of its application varied uniformly within a two-second time interval (Figure 2.1a; see Methods). Before, it was found that the activity of single neurons covaries with the subject's choice. Here, by analyzing pair-wise spike count correlations, we take a population level approach. This allows us to detect a purely internal signal that correlates the population of neurons.

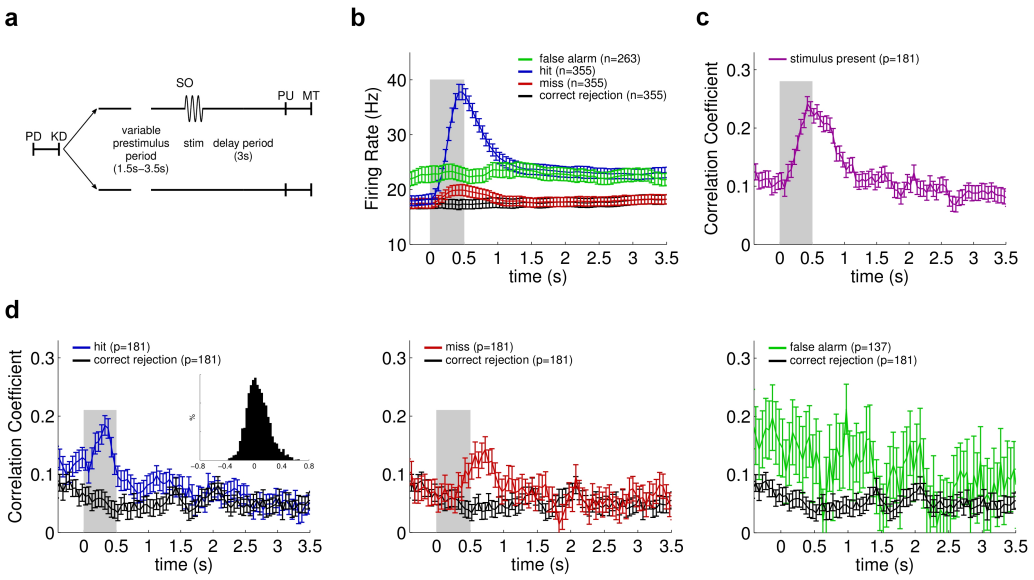

Figure 2.1: **Detection task and temporal profile of the firing rates and the spike-count correlation coefficients (A)** The mechanical probe was lowered indenting the skin of one fingertip of the restrained hand (PD) and the monkey reacted placing its free hand on an immovable key (KD). After a variable pre-stimulus period from 1.5-3.5 s, on half of the trials a vibratory 0.5 s stimulus was presented (SO). After a fixed delay period of 3 s the stimulator probe moved up (PU), indicating to the monkey that it could make the response movement (MT) to one of the two buttons. The button pressed indicated whether or not the monkey felt the stimulus. **(B)** Temporal profile of the firing rates according to the behavioural conditions (n = number of neurons). Stimulus-present trials were aligned at SO and stimulus-absent trials were aligned at PU. Gray box marks the time of stimulus presentation. **(C)**, Temporal profile of the correlation coefficients (CCs) of the spike count using the stimulus-present trials aligned at the SO (p = number of pairs). Gray box marks the time of stimulus presentation. **(D)** Temporal profile of the CCs according to the behavioural conditions. Stimulus-present trials were aligned at SO and stimulus-absent trials were aligned at PU. The time courses of the CCs are modulated in a condition-dependent manner. The inset shows the CCs population histogram of neurons pairs from t=2 to t=3 s from SO for correct reject trials.

2.2 Results

2.2.1 Premotor cortex activity is modulated during the task

The temporal profiles of neuronal firing rates covary with the decision report (de Lafuente and Romo [2005, 2006]) (Figure 2.1b). Interestingly, the firing rate activity during false-alarm trials is higher than during miss and correct-reject trials. This could be an indication that premotor cortex neurons are receiving a stimulus-independent signal. If an internal signal were collectively affecting the neural population, the firing rate of pairs of neurons would cofluctuate, perhaps in a time-dependent way. We have then analyzed the time-course of the spike-count correlations of pairs of simultaneously recorded premotor cortex neurons (de Lafuente and Romo [2005, 2006]). We start by noticing that these noise correlations are modulated during the course of a trial. The temporal profile of the spike-count correlation coefficient (CC), defined in a time window of 250 ms and computed using stimulus present trials aligned at the stimulus onset, is shown in Figure 2.1c. Before stimulus presentation, CCs are relatively weak. Following stimulus onset, and with latency similar to firing rate responses (de Lafuente and Romo [2006]), CCs grow until they are more than twice their initial value.

When trials are segregated according to the animal's decision reports, the time-course of the CCs appears modulated in a condition-dependent manner (Figure 2.1d). One notices that a) the CCs in miss and correct-reject trials have similar temporal profiles, except during the stimulus period and the subsequent relaxation in miss trials; b) in hit trials they reach higher values during the stimulation period; c) the CCs in false-alarm trials are higher than in correct-reject trials during the first half of the shown interval; d) noise correlations can be weak; their smallest values, in the four conditions, are attained during the last portion of the delay period reaching mean values of about 0.06, in agreement with measurements in the supplementary motor area during simple reaching tasks (Averbeck and Lee [2003]). This is seen in the inset of Figure 2.1d showing the distribution of CCs in correct-reject trials at the end of the delay period. Pre-

vious recordings in prefrontal cortex of monkeys performing a working memory task studied the time course of CCs, but correlations were not modulated at any task stage (Constantinidis and Goldman-Rakic [2002]).

2.2.2 Spike-count correlations and firing rate activity reveal a purely internal processing

A closer view to Figure 2.1d shows that before stimulus onset the CCs in hit trials are higher than those obtained for correct rejections at any time. This could be another manifestation of an internal signal that, starting before stimulus onset, modulates the activity of premotor cortex neurons and might influence the outcome of the decision-making process. To further explore the properties of this signal, we computed the CCs by aligning the trials when monkeys place their non-stimulated hand on the immovable key (KD; Figure 2.1a). This event is important because it initiates the 1.5-3.5 s variable period that precedes stimulus onset. We hypothesized that if premotor cortex neurons reflect the animal's use of the knowledge of stimulus onset times, modulation of their CCs should start roughly at 1.5 s. Again, the time course of the CCs depends on the behavioral condition (Figure 2.2a). In stimulus-present trials, we computed the time course of these coefficients keeping trials only until the application of the stimulus. In all decision reports, we observed a modulation that seems to be driven by the internal signal. The temporal profiles of CCs during miss and correct-reject trials are similar (Figure 2.2a). In accordance with the hypothesis that the internal signal influences the decision reports, we observed that during hit and false-alarm trials the CCs undergo a large positive fluctuation beginning about 1.5 s after KD. Similar temporal dynamics can be observed on the firing rates: while in correct-reject and miss trials they become stationary soon after KD, in false-alarm and hit trials they begin to increase at about 1.5 s (Figure 2.2b).

The modulated activity occurring before stimulus onset is consistent with the hypothesis that premotor cortex neurons make use of task contingencies to prepare the network for the stimulus arrival. The higher average firing rate in false-alarm trials beginning from KD suggests the presence of an internal signal

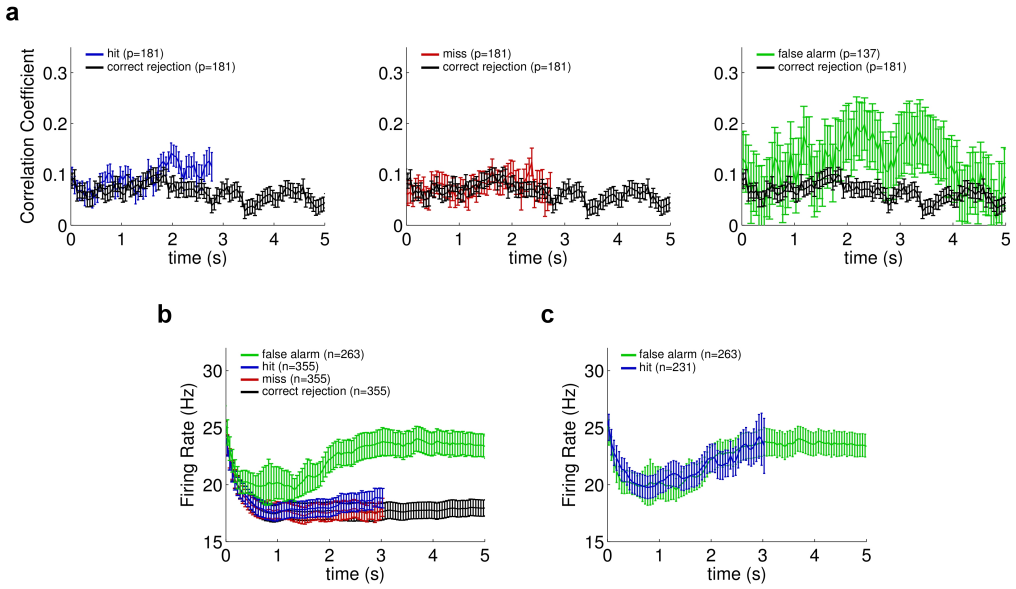

Figure 2.2: **Temporal profile of the spike count correlation coefficients and mean firing rates in trials aligned at key down (KD)**. In all panels trials were segregated according to the behavioural condition and stimulus-present trials were kept only until the application of the stimulus. **(A)** Time course of the CCs for each behavioural condition (p = number of pairs). During hit and false-alarm trials the CCs increase starting approximately 1.5 s after KD. **(B)** Mean firing rate in each behavioural condition (n = number of neurons). As with the CCs, in hit and false-alarm trials, the firing rate increases after 1.5 s immediately after KD. **(C)** Mean firing rate of a subset of hit trials equivalent to the mean firing rate of false alarms. This subset was obtained selecting those subthreshold stimulus amplitude hit trials with mean firing rate higher than 6 Hz.

controlling the excitability of the neurons that fluctuates from trial to trial. Its effect in the other behavioral conditions is less evident presumably because it is weak and therefore the probability that neurons reach their firing rate threshold is low (Figure 2.2b). However, the fluctuating nature of the signal can be made more visible in the hit condition by restricting the computation of firing rates to trials with weak stimulus amplitudes (less than 6 μm) and mean firing rates higher than 6Hz (Figure 2.2c). The internal signal has an appreciable strength

as can be seen by comparing the change in firing rate that occurs in false-alarm trials at 1.5 s after KD (about 5 Hz) with the changes produced by the stimulus in hit trials (about 20 Hz) and in miss trials (about 3 Hz).

2.2.3 The activity of premotor cortex neurons covaries at slow temporal scales

The presence of temporally modulated noise correlations before stimulus onset suggests that the fluctuating signal is common to a substantial number of neurons. To quantify this effect we studied the covariation of pairs of neurons at scales longer than T=500 ms. Slow excitability cofluctuations of specific pairs can be detected by comparing the covariance of the spike-counts of the neurons, defined in a time window of size T, with the product of the two spike-counts. Since the spike-count variable does not have information about fluctuations at scales shorter than T, the remaining covariations are guaranteed to originate from slower scales. On the other hand, the product of the spike counts supports the hypothesis of firing independence at scales longer than T. If two neurons did not cofluctuate at long time scales, these two quantities should be equal; a non-zero value of their difference E reveals the presence of slow covariations (see Methods). This analysis confirms that premotor cortex neurons do covary at time scales longer than 500 ms. To trace the effect of the internal signal on the population of pairs, we computed the distribution of E at two different 500 ms bins of the task during hit trials (before stimulus onset and at the end of the delay period). The result in Figure 2.3a indicates that before stimulus onset many pairs share the internal signal (left panel), but by the end of the delay period the fraction of pairs with a low level of covariation increases substantially (right panel). The long tail exhibited by the distribution in the pre-stimulus period disappears during the delay period, being replaced by a larger peak at the origin (Figure 2.3b).

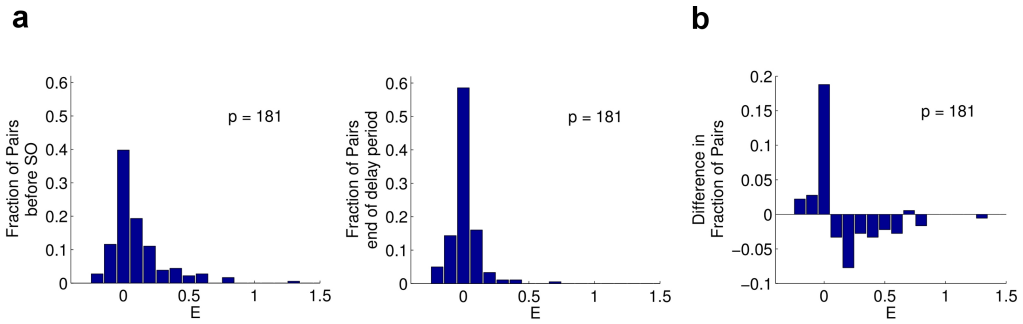

Figure 2.3: **Analysis of slow covariations (A)** Distribution of E for 181 pairs of neurons before SO (top) and at the end of the delay period (bottom). Mean value of E is 0.10 during the prestimulus period, but reduces to 0.01 at the end of the delay period. **(B)** Difference in the fraction of pairs between the prestimulus period and the end of the delay period. Many pairs share the internal signal before the application of the stimulus, but little before the animal reports his decision the fraction of pairs with a low level of covariation increases substantially.

2.3 Discussion

The neuronal fluctuations described above occurring before stimulus onset could be reflecting the initiation of the decision making process. If so, some features of the strategy developed by animals to solve the detection task can be inferred from the neuronal activity preceding stimulus onset. If this is the case, the problem posed by the uncertainty in the task could be solved by applying an internal reference signal at that time. If the strength of this pulse were such that in the absence of stimulation the population activity remained below the decision criterion and the weakest stimulus were large enough to put it just above it, in principle the task could be performed well. However noise spoils this strategy. A possible way to deal with this situation is to take into account the history of the decision reports during previous trials to modify in a flexible way the behavioral response in the current trial, something that could be implemented by modulating the neuronal depolarization at long time scales. In this context, the activity during the pre-stimulus period would be related to the inference about the presence of the stimulus in the current trial based on the recent history. In fact, in this task, the number of yes responses before false-alarm trials is larger

than in the set of all trials (Hernández et al. [2002]), which is an indication that the decision making process uses memory in a time scale longer than one trial.

The internal fluctuating signal could be produced by neuromodulatory systems, which are known to be involved in decision-making tasks (Aston-Jones and Cohen [2005]; de Lafuente and Romo [2011]; Doya [2008]; Yu and Dayan [2005]). In the task studied here, the activity of midbrain dopamine neurons is correlated with the monkeys decision report (de Lafuente and Romo [2011]). However, it does not present important modulations during the pre-stimulus period. The noradrenergic system has been suggested to be related to uncertainty aspects in detection tasks (Aston-Jones and Cohen [2005]; Yu and Dayan [2005]), but experimental studies concluded that this signal is produced after the decision has been made in cortical areas (Aston-Jones and Cohen [2005]). Thus, it is unlikely that this system could play a role in the activity changes observed during the pre-stimulus period. Although the current evidence seems to be against of noradrenaline being responsible for the generation of the fluctuating signal, this conclusion has to be taken with some caution because a task with the same type of uncertainty present in our work has not been studied experimentally.

The activity observed during the pre-stimulus period may result from reverberating activity occurring in a distributed set of prefrontal and premotor areas which have been shown to be involved in working memory (Goldman-Rakic [1995]; Hernández et al. [2002, 2010]; Romo et al. [1999, 2004]), decision making (Hernández et al. [2002, 2010]; Merten and Nieder [2012]; Romo et al. [2004]), stimulus selection and movement preparation (Romo and Schultz [1991]; Schall [2001]; Tanji [2001]).

Accumulator models (Gold and Shadlen [2000]; Hanes and Schall [1996]; Shadlen and Newsome [2001]; Shadlen et al. [1996]) have been successful in explaining some decision-making experiments (Gold and Shadlen [2000]; Shadlen and Newsome [2001]). However, as noted previously in sensory areas (Nienborg and Cumming [2009]), feedforward bottom-up processing cannot fully explain experimental results. In our detection task the time when accumulation of sensory evidence should start is ambiguous, but the moment after which the stimulus could be applied is well-defined and the neural population does initiate integration at that

time (Figure 2.2b,c). However, this integration is effective only if cells are sufficiently depolarized. Thus, the behavioral response could result of a combination between the internal signals and the sensory input; hit responses may result from stimulus-present trials where neurons are highly depolarized (in which case the stimulus is not relevant for the yes response) or from trials where the depolarization is not enough to reach the decision criterion, but the added effect of the stimulus suffices to obtain the correct response.

Decisions are choices made under uncertain conditions (Cook and Maunsell [2002]; de Lafuente and Romo [2005]; Platt and Glimcher [1999]; Schall [2001]; Shadlen and Newsome [2001]; Sheinberg and Logothetis [1997]). Tasks in which sources of uncertainty can be controlled provide excellent conditions where to unveil the internal signals involved in decision-making processes. Noise correlations in cortical networks can be quite small (Ecker et al. [2010]; Renart et al. [2010]) and observation of appreciable modulations in the covariation of pairs of neurons can be a signature of the presence of common internal signals. Our results could be pointing to a role of Bayesian inference in the cortical network where the internal signal reflects the animal's belief about the state of the world (Beck et al. [2008]; Rao [2010]) caused by the uncertainty about the amplitude and application time of the stimulus. Future experimental and theoretical work could clarify the connection between purely internal cortical processing and types of uncertainty in the task.

2.4 Methods

Data for this analysis were obtained from two earlier studies (de Lafuente and Romo [2005, 2006]). Neuronal recordings were obtained with an array of seven independent, movable microelectrodes (2-3 MΩ) inserted in areas VPc, DPc and MPc, bilaterally. A total of 355 neurons were included in the analysis, based on their response to any of the different components of the task and the stability of the recordings. Trials were classified according to monkeys choice and stimulus amplitude in hits, false alarms, misses and correct rejections. For each condi-

tion, neural recordings were used for computing the firing rate and correlation coefficients if they were at least 5 trials.

Firing rate as a function of time was calculated using a 250 ms sliding window displaced every 50 ms. In order to combine sets of trials with different stimulus amplitudes, we subtracted from each trial the mean firing rate of the set of trials with equal amplitude and divided by its standard deviation.Correlation coefficients as a function of time were calculated from the firing rates of each pair of simultaneously recorded neurons following:

$$CC(t_i) = \frac{\langle \nu_1(t_i)\nu_2(t_i)\rangle - \langle \nu_1(t_i)\rangle\langle \nu_2(t_i)\rangle}{\sqrt{\mathrm{var}(\nu_1(t_i))\mathrm{var}(\nu_2(t_i))}} \tag{2.1}$$

where $\nu_k(t_i)$ is the firing rate of a neuron k at window i. Confidence intervals were estimated using a bootstrap technique. In each window we generated 500 re-samples of the firing rates of the pair of neurons with the same number of trials as the original one. Re-samples were drawn from the same collection of trials with which the CC was calculated. From these re-samples we obtained a distribution of correlation coefficients and the confidence interval was considered proportional to the variance of this distribution (significance evel: 90%, two-sided). The mean temporal profile of the correlation coefficient over all pairs was computed using a weighted average. For each window and for each pair, the weight of the correlation coefficient value was considered proportional to the inverse of its confidence interval.

The analysis of slow covariations was done computing the distribution of E over the population of pairs of neurons. We defined E as,

$$E = \frac{1}{N}\frac{\sum n_i^k n_j^k}{\left(\frac{1}{N}\sum_{k=1}^{N} n_i^k\right)\left(\frac{1}{N}\sum_{k=1}^{N} n_j^k\right)} \tag{2.2}$$

where n_i^k and n_j^k are the spike counts of neurons i and j in trial k, computed in a time window of length T=500 ms and N is the number of trials. Following this definition, any deviation from E = 0 indicates a covariation of the spike

counts larger than the expected for independent neurons. Since E is computed from spike counts in time windows of length T, values of E different from zero indicates covariations at time scales longer than T. The distribution of E over the population of pairs was calculated in two 500ms-long periods of the task: before SO from t = 2s to t = 2.5s following KD and end of delay period from t = 3s to t = 3.5s from SO. For the first period trials were only considered if stimulus was presented after t = 2.5s following KD. The histograms were computed using a bin size of 0.1 and normalized with the total number of pairs.

Chapter 3

Decoding decisions from neural populations

"You're nothing but a pack of neurons"
— Francis Crick, *The Astonishing Hypothesis* (1994)

3.1 Introduction

When decisions are based on sensory evidence, decision-related signals evolve across sensory and frontoparietal cortices (for reviews, see Gold and Shadlen [2001]; Romo and de Lafuente [2013]; Romo and Salinas [2003]). The involvement of single neurons in decision-making processes is usually studied in terms of the choice probability (CP) index, a measure of covariation between a neurons firing rate activity and the subjects choice (Britten et al. [1996]; Green et al. [1966]. In the brain, however, decisions engage multiple pools of neurons distributed across brain areas (de Lafuente and Romo [2006]; Hernández et al. [2010]; Romo and de Lafuente [2013]). Hence, if one is to decode behavioral choices, the relevant measurements must come from population variables constructed from the spiking activity of multiple neuronal pools.

To understand how decisions emerge, one must first define proper measures to

quantify how population activity covaries with behavior. It is known that firing rates vary stochastically from trial to trial (Shadlen and Newsome [1998]; Tolhurst et al. [1983]) and that pairs of neurons exhibit correlated variability (Gawne and Richmond [1993]; Zohary et al. [1994]), often named noise correlation. Such correlations between neurons strongly impact the association between neuronal activity and behavior (Shadlen et al. [1996]), in particular, it is known that the CP index depends on the correlation structure of the neuronal network (Cohen and Newsome [2009]; Haefner et al. [2013]; Nienborg and Cumming [2010]; Nienborg et al. [2012]). In addition, we recently demonstrated that the temporal profile of the noise correlation coefficient changes as the task progresses, reflecting dynamic effects of stimuli and internally generated signals on frontal lobe neurons that might participate in the decision process (Carnevale et al. [2012]). Given that a decision evolves over time, we think it is important to detect and describe transient interpool interactions. Some knowledge about the dynamics of a large-scale cortical network during decision making has been obtained by studying macroscopic signals from magnetoencephalographic recordings (Siegel et al. [2011]), but the dynamical profile of correlations has rarely been studied at the circuit level (Pesaran et al. [2008]).

Motivated by these observations, we developed analytical tools to study the dynamics of neuronal pools and their relation to behavior. We tested these tools with data from simultaneous recordings of neuron pairs obtained while monkeys performed a decision-making task (de Lafuente and Romo [2005, 2006]). Specifically, we have first extended the concept of CP index, which traditionally refers to single neurons, to define measures of covariation between behavior and the firing rates of two or more neurons. We have then derived analytical expressions that explicitly relate these measures to statistical properties of the pools' spiking activity, obtaining a precise description of how noise correlations affect the standard CP index and the generalized indices introduced here. We find that the CP becomes significant when the correlation coefficients depend strongly on the choice outcomes of the trials used to compute them and that the association between population activity and behavior increases notably when the choice-conditioned correlations are small.

To address the issue of how neuronal pools cooperate to form the decision, we reasoned that the concerned pools combine their firing outputs and send the resulting signal downstream for further processing. Since the observed behavior is a consequence of these neural computations, we assumed that an important combination of pool activities would be one that covaries closely with behavior.

To test these ideas, we analyzed simultaneous recordings of pairs of premotor cortex (PM) neurons of distinct functional types and also of neuron pairs in the secondary somatosensory area (S2) (de Lafuente and Romo [2005, 2006]). In the decision-making task, monkeys had to detect a stimulus that often was very weak and was absent in half of the trials. Both the PM and S2 areas contained two types of neurons that exhibited oppositely tuned responses to stimulation (de Lafuente and Romo [2006]). Presumably, these two neuronal pools contribute to the decision-making process. For the detection task analyzed here, we found that sensory-like neurons in PM areas covary strongly with the decision report during the stimulation period, although this covariation does not reach its largest possible value. In contrast, pools of PM neurons exhibiting delay activity during the period between the application of the stimulus and a cue signal that triggers the decisions motor report become fully correlated with the subjects choice. Interestingly, this occurs when the population firing rates of the relevant pools are combined optimally, maximizing the generalized measures of covariance with behavior.

3.2 Results

3.2.1 Measures of covariance between behavior and the activity of neural pools

Consider a perceptual decision-making task in which the subject has to decide between two possible choices, A or B. Covariation between the activity of single neurons and the subject's choice is often quantified by the CP index. This quantity represents the average probability with which an external observer could

predict the subject's decision from the activity of a single neuron, using the accrued knowledge of the firing rate distributions computed over trials in which option A or option B was selected. If the neuron responds identically in trials in which the subject chooses A (A trials) and in trials in which it chooses B (B trials), the prediction performance of the external observer is at chance level (CP = 0.5). Conversely, if the firing rate distributions of the neuron in trials A and B are fully distinct, the external observer could perfectly predict the subjects decision (CP = 1).

The CP index can be computed as the area under the receiver-operating characteristic curve (ROC) of the neuron's firing rate, segregating trials according to the subject's choice (Britten et al. [1996]; Green et al. [1966]). If the neuron's firing rate distributions in trials A and B can be described as Gaussian, one finds the following analytical expression (see Appendix A for the derivation)

$$\mathrm{CP} = \frac{1}{2}\mathrm{erfc}\left(-\frac{\delta}{2}\right) \qquad \delta = \frac{\mu^A - \mu^B}{\sqrt{\frac{1}{2}\left(\sigma_A^2 + \sigma_B^2\right)}} \tag{3.1}$$

where μ^c and σ_c^2 are the mean and variance of the firing rate over trials in which the subject's choice was c = A, B. The quantity δ is the difference between the firing rate means in trials A and B, measured in units of the arithmetic mean of the two variances.

The CP index is a useful measure to study how the activity of a single neuron covaries with behavior. However, the decision-making process is determined by neuronal populations (de Lafuente and Romo [2006]; Heekeren et al. [2004]; Hernández et al. [2010]; Pesaran et al. [2008]; Siegel et al. [2011]). Understanding how the decision is formed in the brain requires the use of proper measures to quantify the covariance of population activity variables with the subject's choices. This can be done by extending the concept of CP index to the combined activity of several neurons. Here we start by considering the case in which cells can be sorted into homogeneous pools of similar responses, and in Section Finding the optimal decision code, we study the general case. For the simplest example, two

neurons from the same pool, we consider the ROC index of the sum of their firing rates r_1 and r_2 , $r_w = r_1 + r_2$, which can be estimated by

$$\mathrm{CP}_{2,w} = \frac{1}{2}\mathrm{erfc}\left(-\frac{\Delta_w}{2}\right), \quad \Delta_w = \frac{\mu_w^A - \mu_w^B}{\sqrt{\frac{1}{2}\left(\sigma_{w,A}^2 + \sigma_{w,B}^2\right)}} \tag{3.2}$$

where μ_c and $\sigma_{w,c}^2$ are the mean and variance of r_w over trials of choice c. This can be expressed in terms of the firing properties of the pair of neurons

$$\Delta_w = \frac{\sqrt{2}\delta_{1,2}}{\sqrt{1+\rho_{12}^w}} \tag{3.3}$$

Here $\delta_{1,2}$ is the arithmetic mean of δ_1 and δ_2 , defined in Equation 3.1, and $\rho_{12}^w = 0.5\left(\rho_{12}^{w,A} + \rho_{12}^{w,B}\right)$ is the arithmetic mean of the correlation coefficients between r_1 and r_2 , computed over trials of decision A and B, $\rho_{12}^{w,A}$ and $\rho_{12}^{w,B}$ (see Equation 3.11 in Methods). The superscript w indicates that the two neurons belong to the same pool. For simplicity, in Equation 3.3 we assumed that the variance of the single neuron's firing rate distributions is equal for both neurons, in both trial types (A and B). Equation 3.3 relates the subject's choices to the activity of the pool of neurons, but it does so in terms of the properties of the two neurons in the pool and their interaction, as captured by the correlation coefficient. The general expression is given in the Appendix A (Equation A.16). Notice that $\mathrm{CP}_{2,w}$ will always be higher than CP, except when $\rho_{12}^w = 1$. This is a reasonable result: the averaged activity of two neurons in the same pool covaries with behavior more than that of single neurons, provided that their responses are significantly different, i.e., that their correlation is not too large.

For two neurons in different pools, we consider an arbitrary linear combination of their firing rates, $r_b = C_1 r_1 + C_2 r_2$, and quantify its covariation with the subject's choices by another ROC index, $\mathrm{CP}_{2,b}$. This index can be estimated as

(see Appendix A for details)

$$\mathrm{CP}_{2,b} = \frac{1}{2}\mathrm{erfc}\left(-\frac{\Delta_b}{2}\right), \quad \Delta_b = \frac{\delta_1 + D\delta_2}{\sqrt{1 + D^2 + 2D\rho_{12}^b}} \tag{3.4}$$

where $D = C_2/C_1$. Here $\rho_{12}^b = 0.5\left(\rho_{12}^{b,A} + \rho_{12}^{b,B}\right)$ with $\rho_{12}^{b,c}$ being the correlation coefficient between the firing rates of neurons 1 and 2 in different pools, computed over trials of choice c. Again, we assumed that the variance of the firing rates is equal for both neurons and both types of trials (see general expression in Equation A.22). Note that if $\rho_{12}^b < 0$ and $D < 0$, the $\mathrm{CP}_{2,b}$ index increases as $|\rho_{12}^b|$ decreases.

To test the amount of covariation with behavior of the combined activity of different neural populations, we consider a further extension of this procedure. Given a set of P pools each having N neurons and a population firing rate r_α ($\alpha = 1, ..., P$), we can quantify the amount of covariation of an arbitrary combination of the pools firing rates, $r_N = \sum C_\alpha r_\alpha$, defining the $\mathrm{CP_N}$ index,

$$\mathrm{CP_N} = \frac{1}{2}\mathrm{erfc}\left(-\frac{\Delta_N}{2}\right) \quad \Delta_N = \frac{\mu_N^A - \mu_N^B}{\sqrt{\frac{1}{2}\left(\sigma_{N,A}^2 + \sigma_{N,B}^2\right)}} \tag{3.5}$$

In applying these calculations to experimental data, we will be interested in linear combinations of rates from two pools, $r_N = C_1 r_+ + C_2 r_-$, with the pools defined as + and −. Δ_N can be expressed in terms of population-averaged firing rates, variances, and correlation coefficients,

$$\Delta_N = \frac{\sqrt{N}\left(\bar{\delta}_+ + D\bar{\delta}_-\right)}{\sqrt{\left(1 + D^2\right)\left[1 + \left(N - 1\right)\bar{\rho}^w\right] + 2DN\bar{\rho}^b}} \tag{3.6}$$

where again $D = C_2/C_1$. The bar indicates population average. For simplicity,

we took equal population averaged variances for the two pools and the two choices (see the general expression in Equation A.31. Correlation coefficients $\bar{\rho}^w$ and $\bar{\rho}^b$ affect the $\mathrm{CP_N}$ index in a manner similar to that for $\mathrm{CP}_{2,w}$ and $\mathrm{CP}_{2,b}$, respectively. The factor that amplifies the population-averaged δ's increases as $\bar{\rho}^w$ decreases. For $D < 0$ and $\bar{\rho}^b < 0$, Δ_N also increases when $|\bar{\rho}^b|$ decreases. For the particular case in which $(N-1)\,\bar{\rho}^w >> 1$, Δ_N becomes independent of the number of neurons. If in addition, $\bar{\rho}^b \sim 0$, $\bar{\rho}^w$ still modulates the amplification of this index, $\Delta_N \propto (\bar{\rho}^w)^{-\frac{1}{2}}$.

3.2.2 Finding the optimal decision code

The measures defined above can be used to study the interaction of neural pools during the decision-making process. If two pools cooperate in forming the decision, then combinations of their firing rates must covary with the behavioral response. But the reverse should also be true: maximizing this covariation should lead to the combination of rates that optimally predicts the animal's decision. This can be done by optimizing the $\mathrm{CP_N}$ index with respect to the relative contribution of the two pools to the population variable (D), which is equivalent to maximizing the mean difference between choices divided by the choice-conditioned variances (Equation 3.5).

To study how population activity covaries with behavior, we considered the case in which cells can be assigned to one of two discrete pools. If neurons could be sorted into discrete and homogeneous pools, one would assign equal weights to all neurons within the same pool. However, the assumption of neurons distributed in discrete pools can be relaxed. In a more general case, neurons contribute to the population variable in a graded manner, with their firing rates weighted with different coefficients. A $\mathrm{CP_N}$ index associated with this variable can still be defined and expressed in terms of its means, variances, and covariance matrix in the two conditions (see Appendix A). The proposed optimality criterion is again equivalent to maximizing the mean difference between choices divided by the choice-conditioned variances. Equivalently, finding the weights amounts to obtaining the Linear Fishers Discriminant between the two decisions (Equations

A.37 and A.38).

In this more general situation, the CP_N index depends on the covariance between every pair in the population. Thus, searching for optimal population variables in experimental data requires the simultaneous recording of multiple neurons in the population. However, in situations in which neurons can be classified into discrete pools, the CP_N can be computed from Equations 3.5 and 3.6. The degrees of freedom involved in the optimization procedure are reduced to one: the coefficient D that combines the pools' activities. Besides this parameter, in this case, the CP_N index only depends on two population-averaged correlation coefficients, one for neural pairs within each pool and another for pairs between pools.

Notice that the optimization procedure does not assume the existence of a decision rule based on the neurons' firing activity. In our formalism, the population variable could be any combination of the firing rates and is not necessarily related to a decision rule. The proposed procedure can be seen as a way to search for population variables that are optimally correlated with behavior. It can be applied to neurons in any area participating in the decision-making process.

3.2.3 The covariance between global activity and behavior is determined by the network correlation structure

We now turn to a more detailed analysis of how the correlation structure affects the covariance between firing activity and behavior. The CP index in Equation 3.1 is computed from properties of single neurons (means and variances of the firing rate distributions in trials A and B). Although pairwise correlations do not appear explicitly in this equation, the CP index does depend on the correlation structure of the neural population involved in the decision-making process. This is because the firing rate distributions are conditioned to the subjects choice, which is determined by the network state during the trial.

To make this dependence explicit, we must relate the usual correlation coefficient R_{ij} for the neuron pair (i,j) (that is, the correlation coefficient computed

using all trial types, regardless of the subject's choice) with the difference in mean firing rates between trials ending in each of the two choices. The latter are essentially the quantities δ, defined for neurons i and j, as in Equation 3.1. Given an arbitrary partition of the set trials into two different groups (A and B), the correlation coefficient R_{ij} can be expressed as

$$R_{ij} = \frac{\frac{1}{2}\left(\rho_{ij}^{A} + \rho_{ij}^{B}\right) + \frac{1}{4}\delta_i\delta_j}{\sqrt{\left[1 + \left(\frac{\delta_i}{2}\right)^2\right]}\sqrt{\left[1 + \left(\frac{\delta_i}{2}\right)^2\right]}} \tag{3.7}$$

where, for simplicity, we assumed that the variance of the firing rates of the two neurons is equal in both types of trials (see Equation A.48 for the general expression). Equation 3.7 shows that, apart from a common factor, the correlation coefficient R_{ij} is the sum of two effects: a contribution from the difference in rates between the two choice conditions (δ_i and δ_j , see Equation 3.1) and a contribution from the choice-conditioned correlation coefficients, ρ_{ij}^{A} and ρ_{ij}^{A}.

If the network contains several neural pools, the correlation structure consists of correlation coefficients of pairs of neurons in the same pool and correlations between neurons in different pools. Given a triplet of cells (1, 2, 3), with neurons (1, 2) in the same pool and neuron 3 in a different pool, we can use the equation above for the pairs (1,2), (1,3), and (2,3). This leads to equations for R_{12}, R_{13}, and R_{23} that can be solved for δ_1, δ_2, and δ_3, obtaining the CP indices as a function of correlations between neurons in the same or different pools. However, implementing this procedure to analyze electrophysiological data requires the simultaneous recording of triplets of neurons. Moreover, the mathematical solution in terms of correlation coefficients between the three pairs of neurons becomes rather cumbersome. It is desirable to have a way to estimate the CP index using only data from simultaneous recordings of pairs of neurons. Now we show that it is possible to obtain a simple approximate expression for the population-averaged CP index based only on correlations between pairs of neurons. It is enough to assume that, given two neurons (1,2) in the same population, $\delta_1 \sim \delta_2 \sim \delta_0$. Using

R_{12} we obtain

$$\mathrm{CP} \sim \frac{1}{2}\mathrm{erfc}\left(-\frac{\delta_0}{2}\right), \quad \left(-\frac{\delta_0}{2}\right) \sim \frac{R_{12}^w - \rho_{12}^w}{1 - R_{12}^w} \tag{3.8}$$

where $\delta_0 = (\delta_1 + \delta_2)$ (see Equations A.50-A.53 in the Appendix A for a discussion of the accuracy of this approximation). Averaging Equation 3.8 over the population of independent pairs gives the estimate for the population-averaged CP index. This is our main result. It shows that the CP index is different from 0.5 when correlations evaluated using all trial types differ from the correlations conditioned on the subjects choice. Neurons could covary significantly with behavior even if the latter correlations are very small. It has been pointed out that correlated activity is necessary for observing robust covariations between single neuron responses and behavior (Shadlen et al. [1996]). On the other hand, pairwise correlations in recurrent networks can be quite small (Ecker et al. [2010]; Renart et al. [2010]). The equations above show that there is no contradiction between these two statements: decorrelation in the recurrent network makes rho_{12}^w small, but there is still a contribution to CP coming from R_{12}^w , which is produced by the difference in firing rates between trials A and B (Equation 3.7; see also Brody [1999]). In fact, Equation 3.8 shows that the CP index is maximized when the overall correlations R_{12}^w are large but the choice-conditioned correlations are small. Notice that Equation 3.8 does not assume any model that mechanistically relates the activity of the neurons to the subjects decision.

We can use Equation 3.8 to draw several conclusions. First, notice that $R_{12}^w - \rho_{12}^w \geq 0$: correlations for neurons in the same pool, computed with fixed-choice trials, are smaller than those obtained with the whole set of trials. Instead, if neurons (1,3) in two different pools have opposite mean responses in the two conditions: $\mu_1^A - \mu_1^B \sim \left(\mu_3^A - \mu_3^B\right)$ (that is, $\delta_1\delta_3 < 0$), from Equation 3.7, we observe that $R_{13}^b - \rho_{13}^b \leq 0$ (see Appendix A). In both cases, the sign is determined by the difference in the mean activities in the two trial types. Finally, neurons in a given pool show CP = 0.5 if pairwise correlations in that pool obey $R_{12}^w = \rho_{12}^w$.

Similar considerations apply to the other generalized choice probability indices (Equations 3.3, 3.4, and 3.5). In particular for the CP_N index (Equations 3.5 and 3.6), apart from the population-averaged δ_+ and δ_- , for which the discussion above still holds, there is a factor depending only on the choice-conditioned correlation coefficients. A potential effect of this factor is to amplify Δ_N, thereby pushing CP_N to saturation, that is, closer to full covariance between firing activity and behavior.

Equation 3.8 was obtained for a pair (1,2) of neurons in the same pool. As a more complex example, we now consider a two-pool network satisfying the condition that for a pair (1,3) of neurons in different pools, $\delta_3 \sim -\delta_1 \sim -\delta_0$. This can be seen as a constraint on the correlation structure of the network. Using this constraint, and replacing pairwise correlations by their population-averaged values ($\bar{R}^w$, $\bar{R}^b$, $\bar{\rho}^w$, $\bar{\rho}^b$), it is interesting to observe that the average CP index can be estimated as

$$\mathrm{CP} \sim \frac{1}{2}\mathrm{erfc}\left(-\sqrt{\frac{\left(\bar{R}^w - \bar{R}^b\right) - \left(\bar{\rho}^w - \bar{\rho}^b\right)}{2 - \left(\bar{R}^w - \bar{R}^b\right)}}\right) \tag{3.9}$$

This equation shows an explicit dependence on the difference $\bar{R}^w - \bar{R}^b$ but not on $\bar{R}^w + \bar{R}^b$. In a computational model constrained as in this example, where $\bar{R}^w$ and $\bar{R}^b$ were considered to be free parameters, it was found that the CP index depends only on $\bar{R}^w - \bar{R}^b$ (Nienborg and Cumming [2010]; Nienborg et al. [2012]). However, that model used an explicit decision rule. In contrast, Equation 3.9 does not make any hypotheses about how the decision is made. To compare our prediction with the modeling results, we have determined the choice-conditioned noise correlations by simulating the same model. We found that $\bar{\rho}^w - \bar{\rho}^b$ depends only on $\bar{R}^w - \bar{R}^b$, while $\bar{\rho}^w + \bar{\rho}^b$ depends only on $\bar{R}^w - \bar{R}^b$ (data not shown), confirming the conclusion reached by Nienborg and Cumming [2010].

3.2.4 Analysis of electrophysiological data from a vibrotactile detection task

In the remaining sections of the paper, we analyze electrophysiological data recorded in a vibrotactile detection task (Figure 3.1, see Methods), using the analytical results derived above. We analyze data from S2 and PM. Previous studies of these data showed that in both areas neuronal activity covaries with the animals behavior (de Lafuente and Romo [2005, 2006]). Importantly, it was found that this covariation is related to the animals perception of the sensory stimulus rather than to the selection of the motor plan. Trials were classified as hits (H), misses (M), correct rejections (CR) or false alarms (FA), depending on whether the stimulus was present or absent and on the behavioral response (Figure 3.1B). We analyzed stimulus-present trials, so types A and B (as denoted in all the equations above) correspond to H and M, respectively.

PM neurons were classified according to their responses to a strong stimulus in H trials (de Lafuente and Romo [2006]) (Figure 3.1C). Those that responded only during the stimulation period were labeled as sensory-like neurons and those showing sustained activity during the delay period were classified as delay-activity neurons. In addition, both S2 and PM neurons were labeled as positive if their firing rate transiently increased with the stimulus and as negative if their firing rate decreased in response to the stimulus. These criteria define two oppositely tuned neuronal pools (denoted as positive and negative) for each of the three populations of neurons (S2, sensory-like PM, and delay-activity PM).

We start by showing that employing two neurons to predict the animals choice increases the level of covariation with behavior. We considered two neurons from the same neural population and compared the $CP_{2,w}$ index with the pairwise averaged CP. Figure 3.2 shows the temporal evolution of these two quantities for the population of positive sensory-like PM neurons (Figure 3.2A), positive delay-activity PM neurons (Figure 3.2B), and positive S2 neurons (Figure 3.2C). As expected, the sum of activities of two simultaneously recorded neurons from the same population is more predictive of the animals choice than the activity of single neurons. For S2 (Figure 3.2C) and PM sensory-like neurons (Figure

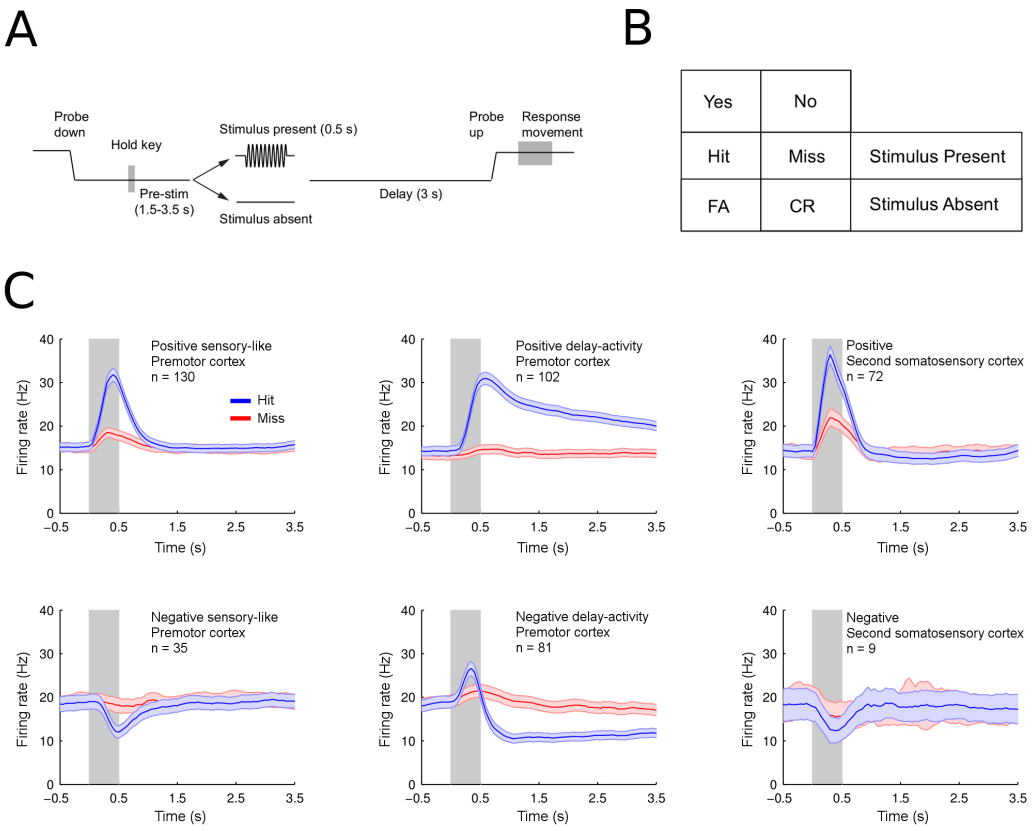

Figure 3.1: **Detection task and neural populations. (A)** The mechanical probe indented the skin of one fingertip of the restrained hand (Probe down) and the monkey reacted by placing its free hand on an immovable key (Hold key). After a variable prestimulus period ($1.5 - 3.5s$), a vibratory 0.5 s stimulus was presented on half of the trials. At the end of a fixed delay period, the stimulator probe moved up (Probe up), instructing the monkey to make a response movement to one of two push buttons. The pressed button indicated whether or not the monkey felt the stimulus. **(B)** A trial is classified according to stimulus presence or absence and to the subjects response as a hit (H), miss (M), correct rejection (CR), or false alarm (FA). Stimulus amplitude was pseudorandomly chosen. A run was composed of 90 trials (amplitude 0) and 90 stimulus-present trials, with varying amplitudes (nine amplitudes with ten repetitions each; $2.3 - 4.6mm$). **(C)** Temporal profile of mean firing rates during hit (blue traces) and miss (red traces) trials for PM neurons (first two columns) and S2 neurons (third column). Top row shows pools of positively tuned neurons and bottom row negatively tuned neurons (n is the number of neurons). Colored area represents SEM.

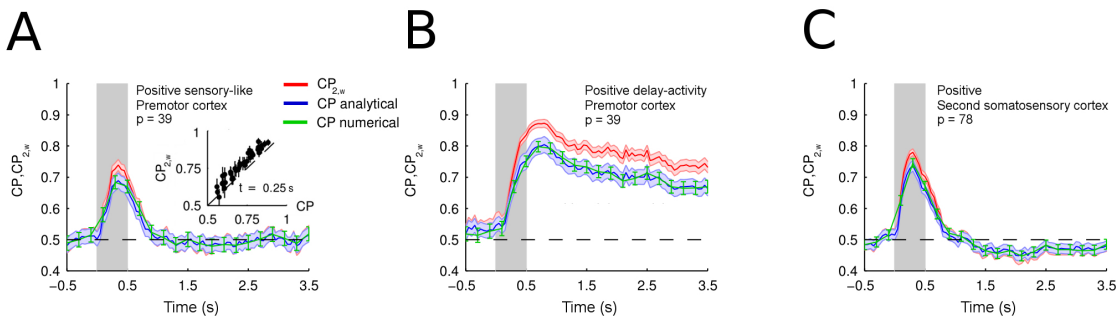

Figure 3.2: **Choice probability obtained by summing pairs of neurons** Temporal profile of population-averaged $CP_{2,w}$ index for pairs of positively tuned neurons (red traces) compared with population-averaged CP index for the same neural pool (blue and green traces). As expected, the combined activity of two neurons better predicts the animal's choice than the activity of a single neuron. **(A)** Pool of positive PM sensory-like neurons. Inset shows $CP_{2,w}$ versus the pairwise averaged CP for each pair ($t = 0.250s$). **(B)** Pool of positive PM delay-activity neurons. **(C)** Pool of positive S2 neurons. $CP_{2,w}$ was obtained from Equation A.16 (see Appendix A). A good agreement can be observed between the analytical CP (Equation 3.1) and its direct numerical evaluation (green traces, see Methods). Gray boxes indicate the period of stimulation; error bars and colored areas represent SEM and p the number of neuron pairs (see also Figure A.1).

3.2A), this is true during the period of stimulus presentation, while for the PM delay-activity neurons, the effect is maintained until the end of the delay period (Figure 3.2B). In addition, the CP indices obtained from Equation 3.1, which assumes Gaussian distributions of responses and involves only their means and variances, are in good agreement with those obtained by direct evaluation, for which no assumption about the response distributions is made (Figure 3.2, blue versus green traces). The $CP_{2,w}$ calculated based on the means and variances of individual neurons and their pairwise correlations (Equation A.16 also compares well with its direct numerical evaluation (Figure A.1).

This analysis shows that these measures of covariance with behavior can be evaluated accurately using only first- and second-order statistics of the neuronal firing rate activity. The analytical expressions can then be used reliably for studying the more general effects of neuronal correlations in the detection task.

3.2.5 How correlated variability determines choice probability

The observation that single neurons covary with the subject's response is usually explained by the existence of correlated variability among the cells in the neuronal population (Shadlen et al. [1996]). This argument refers to noise correlations evaluated over the whole set of trials, a quantity that may receive a substantial contribution from the difference in the firing rates in trials of different choices. However, one may wonder whether noise correlations estimated using subsets of similar trialspresumably with similar firing rates might affect choice probability (or, more generally, any of the indices defined above). Indeed, the results obtained in Equations 3.7 and 3.8 indicate that noise correlation decreases when conditioned on the choice and that these choice-conditioned noise correlations could reduce the CP index.

To investigate this issue further, we start by analyzing the relationship between the correlation coefficients and the difference between the mean firing rates in H and M trials (Equation 3.7). We present this analysis for positive and nega-

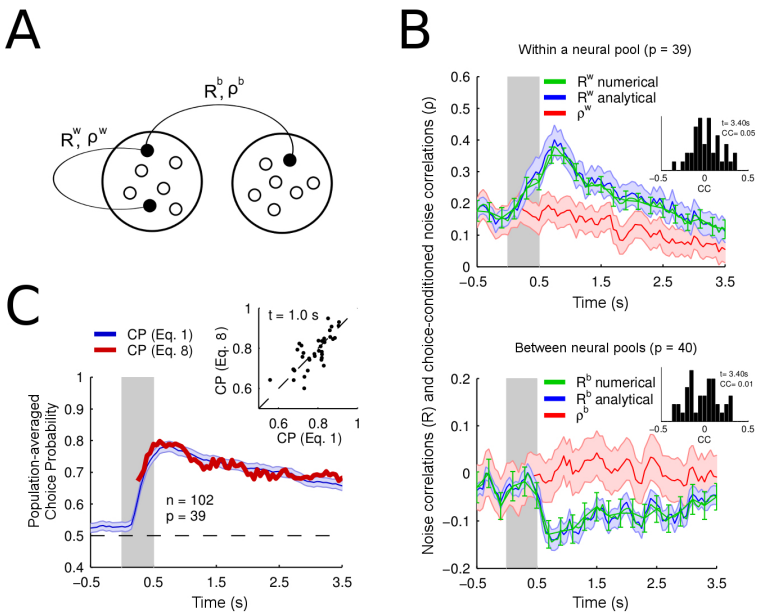

Figure 3.3: **Noise correlations and choice-conditioned noise norrelations determine the CP index. (A)** Correlation structure in a two-pool network. R^w, R^b, ρ^w, and ρ^b denote spike-count noise correlation coefficients between neurons in the same (w) or in different (b) pools. R^w and R^b are computed using all stimulus-present trials, while ρ^w and ρ^b are obtained from trials with a fixed subject's choice.Together, they define the correlation structure of this example network. **(B)** Temporal evolution of mean correlation coefficients of delay-activity PM neurons computed with all trials (R, blue and green traces) compared with average correlations obtained from hit and miss trials separately (ρ, red traces). Top: pairs within the pool of positive delay-activity PM neurons. Bottom: pairs of positive and negative delay-activity PM neurons. Mean correlation coefficients were obtained by averaging over all pairs from the same functional type. Gray boxes indicate the period of stimulus presentation; error bars and colored areas represent SEM and p the number of pairs. Green traces depict the correlation coefficients computed numerically. Blue traces show predictions from Equation A.48. Insets show the distribution of ρ over the population of pairs for a 250 ms time window centered at $t = 3.40s$. **(C)** The CP index of delay-activity neurons computed from correlation coefficients (Equation 3.8) is compared with its evaluation from the mean and variance of the firing rate, in H and M trials (Equation 3.1). Inset shows the pairwise averaged CP computed with Equation 3.1, compared with the CP obtained using Equation 3.8, for each neuronal pair at $t = 1.0s$ (see also Figure A.2).

tive delay-activity neurons. We consider correlation coefficients from pairs within the same neural pool, denoted by the superscript w (that is, R^w and ρ^w), and between neurons from different neural pools, denoted by b (R^b and ρ^b) (see Figure 3.3A). First, we show in Figure 3.3B the temporal profile of the population-averaged correlation coefficients R^w and R^b , obtained by direct numerical evaluation from all trial types (green traces). As we have seen, two separate factors contribute to these correlations: (1) the differences in firing rates in H and M trials and (2) the correlations conditioned on the choice (Equation 3.7). We then compared the same correlations with those obtained analytically by combining these two factors (blue traces, Equation A.48). This comparison shows quite similar values, both for pairs of neurons within the same pool (Figure 3.3B, top) and for pairs of neurons belonging to different pools (Figure 3.3B, bottom, and Figures A.2A and A.2B for pools of S2 and sensory-like PM neurons).

Second, we compared the population-averaged correlation coefficients computed with all stimulus-present trials ($\bar{R}^w$ and $\bar{R}^b$) with those obtained using trials with a fixed choice ($\bar{\rho}^w$ and $\bar{\rho}^b$, red traces). Again, the comparison appears in Figure 3.3B (top) for pairs of neurons within the same pool ($\bar{R}^w$ and $\bar{\rho}^w$) and in Figure 3.3B (bottom) for pairs of neurons from different pools b ($\bar{R}^b$ and $\bar{\rho}^w$). Noise correlations decrease when they are conditioned on the animal's choice: $\bar{R}^w$ exceeds $\bar{\rho}^w$ after the stimulus onset and during the entire delay period. This is explained by our analytic expression, Equation 3.8: for each pair of neurons, the difference $\bar{R}^w - \bar{\rho}^w$ is positive and comes from the difference in mean firing rate in trials of different choice (δ_0). In contrast, when the cells belong to different pools, $\bar{R}^b$ is lower than $\bar{\rho}^b$ (Figure 3.3B, bottom). This is because the firing rate of positively tuned neurons in H trials is larger than in M trials, while the opposite occurs for negatively tuned neurons (Figure 3.1C).

Notice the rather different temporal profiles of correlations conditioned on the choice and correlations defined over the whole set of trials. Whereas the latter are strongly modulated by the stimulus, the choice-conditioned correlations $\bar{\rho}^w$ and $\bar{\rho}^w$ are not. Only toward the end of the delay period does $\bar{\rho}^w$ decrease significantly below the value that it had before stimulus onset (Figure 3.3B, top), although during this period the firing rate of PM delay-activity neurons is higher than

before stimulus presentation. This behavior is consistent with the existence of a common, slowly fluctuating signal that correlates neurons in the positive pool and is present before stimulus onset (Carnevale et al. [2012]). After stimulus offset, this signal tends to disappear and choice-conditioned correlations thus fall to their smallest value at the end of the delay period. In contrast, correlations evaluated using H and M trials remain high because firing rates in these two trial types are different throughout the delay period (top middle panel in Figure 3.1C). Correlations between neurons in different pools, $\bar{\rho}^b$, are much smaller than those between neurons in the same pool.

The difference in the temporal profiles of the correlation coefficients $\bar{R}^w$ and $\bar{\rho}^w$ fully explains the temporal evolution of the population-averaged CP index. In fact, we have seen that this index can be approximated in terms of that difference (Equation 3.8). We studied this prediction using data from the population of PM delay-activity neurons. The average error introduced by this approximation in our data is 11%. The population-averaged CP index, evaluated using only correlation coefficients, is shown in Figure 3.3C, together with the prediction from Equation 3.1. This result confirms that the increase of the population-averaged CP index occurring after stimulus presentation and its subsequent slight decrease during the delay period (Figure 3.2B) are controlled by the transient modulations of the difference $\bar{R}^w - \bar{\rho}^w$ (Figure 3.3B,top). Although mean choice-conditioned correlations can be rather small ($\bar{\rho}^w$ and $\bar{\rho}^b$ in Figure 3.3B, insets), the population-averaged CP index can be large (about 0.7 for this example) because of the contribution from $\bar{R}^w$ (Shadlen et al. [1996]). In fact, correlations conditioned on the choice tend to decrease the covariation of single neurons and of neuronal populations with behavior (Equation 3.8). The diminishing value of these correlations during the delay period helps to maintain a large CP until the subject makes a movement. Further tests of the validity of our analytical results are shown in Figures A.2C and A.2D.

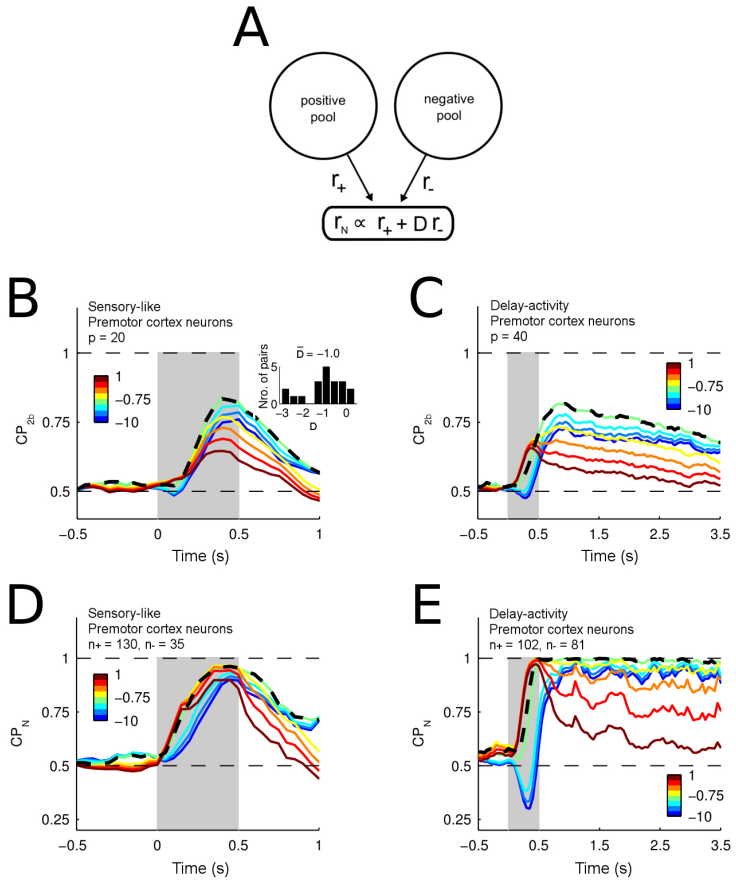

Figure 3.4: **Linear combinations of positive and negative neurons and its covariation with behavior (A)** The activity of a positive pool is linearly combined with the activity of a negative pool. **(B)** Population-averaged $\mathrm{CP}_{2,b}$ for pairs of positive and negative sensory-like neurons, computed from Equation A.22 for different values of D. Color code corresponds to the value of D. Inset shows the distribution of D that maximizes $\mathrm{CP}_{2,b}$ in a $250ms$ window centered at the stimulation period. Dashed line indicates the population-averaged $\mathrm{CP}_{2,b}$ for the mean value of optimal coefficients $D = -1$. **(C)** Same as **(B)** for pairs of delay-activity neurons. In this case, the optimal value of D was obtained by averaging over the second half of the delay period resulting in $D = -1$. Dashed line indicates the population-averaged $\mathrm{CP}_{2,b}$ for this value of D. **(D)** CP_N for the population of sensory-like PM neurons, computed using Equation A.31 for different values of D (color coded). The dashed line corresponds to the CP_N index for the optimal value of D in a 250 ms window centered at the stimulation period, $D = -1.2$. **(E)** Same as panel **(D)** but for delay activity PM neurons. The dashed line corresponds to the CP_N index for the optimal value $D = -0.5$, obtained averaging over the second half of the delay period (as it is explained in Figure 3.5). The number of neurons in the positive and negative pool is denoted by n_+ and n_- respectively. The gray box indicates the period of stimulus presentation (see also Figure A.3).

3.2.6 Choice is unambiguously decoded from the activity of premotor cortex neurons

How well can the population activity of PM neurons predict the subject's choice? How are choices affected by noise correlations? To answer these questions, we considered readout neurons implemented as linear combinations of the outputs of two neural pools (Figure 3.4A). As mentioned before, we reasoned that when two neural pools cooperate to form the decision, the output signal resulting from their interaction must be the one that optimally predicts the subject's decision report. We now use this idea to investigate the involvement in decision making of two neural populations present in PM: sensory-like neurons and delay-activity neurons. Given that these two neuron types presumably play different roles in the process, we have considered them separately. Each of these two populations includes two different pools (+ and-), according to their response to strong stimuli (de Lafuente and Romo [2006]) (Figure 3.1C). Hence, for each population (sensory-like or delay-activity neurons), we have optimally combined the firing outputs of neurons taken from oppositely tuned pools and have analyzed how well this linear combination predicts the decision response.

As a first example, we considered only two neurons from the same population but different pools and we linearly combined their firing outputs. Since the neurons are in different pools, the relevant measure of covariation with behavior is the $CP_{2,b}$ index. Figure 3.4 shows the results for sensory-like PM neurons (Figures 3.4B and 3.4D) and delay-activity PM neurons (Figures 3.4C and 3.4E) for different linear combinations of firing rates (characterized by the coefficient D, color coded). These analyses were much more limited for S2 given our experimental database (see Figure A.3). For pairs of sensory-like PM neurons, the largest values of $CP_{2,b}$ occur during the presentation of the stimulus. The inset in Figure 3.4B shows the distribution of D values that maximize $CP_{2,b}$ in a 250 ms window centered at the stimulation period. The mean value of D over the population of pairs is D = -1, which corresponds to the difference of firing rates between the oppositely tuned pairs of neurons. The dashed line in Figure 3.4B represents the population-averaged $CP_{2,b}$ for this value of D. For delay-activity neurons,

large values of $CP_{2,b}$ are observed during the entire delay period (Figure 3.4C). The most predictive combination of firing rates was again close to the difference ($D = -1$, averaged over the second half of the delay period) and remained constant until the end of the delay period. Again, the dashed line corresponds to the population averaged $CP_{2,b}$ for $D = -1$. These results indicate that, as one might intuitively suspect, the perceptual decision about stimulus presence depends on the difference in activity between the responses of oppositely tuned neurons, in agreement with what has been reported in other perceptual decision-making tasks (Gold and Shadlen [2001]; Heekeren et al. [2004]; Romo and de Lafuente [2013]; Romo and Salinas [2003]). Although in all the above cases the $CP_{2,b}$ index reaches quite large values, it is still well below its largest possible value. Furthermore, the decision-making process probably involves interactions between pools of multiple neurons (Figure 3.4A). Hence, we used the $\mathrm{CP_N}$ index (Equation 3.5 to look for linear combinations of mean firing rates of multiple neurons in oppositely tuned pools that would covary maximally with behavior. For sensory-like PM neurons, the optimal combination is obtained during the stimulation period, with $D = -1.2$ (Figure 3.4D). The dashed line corresponds to $\mathrm{CP_N}$ for this value of D. Although $\mathrm{CP_N}$ is larger than $CP_{2,b}$, it remains below 1 and starts to decrease by the end of the stimulus presentation period.

Most remarkably, the combination of pools of delay-activity PM neurons reaches the value $\mathrm{CP_N} = 1$ soon after stimulus onset and maintains it during the entire delay period (Figure 3.4E). Figure 3.5 (top) shows the temporal profile of the value of D that maximizes the $\mathrm{CP_N}$ index at each time window. This optimal value was obtained independently for each shifted time window, as is illustrated in the inset. After a transient modulation, the optimal value of D becomes stationary with a temporal mean of -0.5 until the end of the delay period. Note that, because D in this case depends on the numbers of neurons in each pool (see Equation A.32), its temporal modulation is important but its specific value is not necessarily so. The transient positive values during the stimulation period are due to the increased activity of the negative pool at that time (Figure 3.1, middle bottom), which produces a transient positive difference between the firing rates in hit and miss trials, opposite to the decrease in activity that this pool ex-

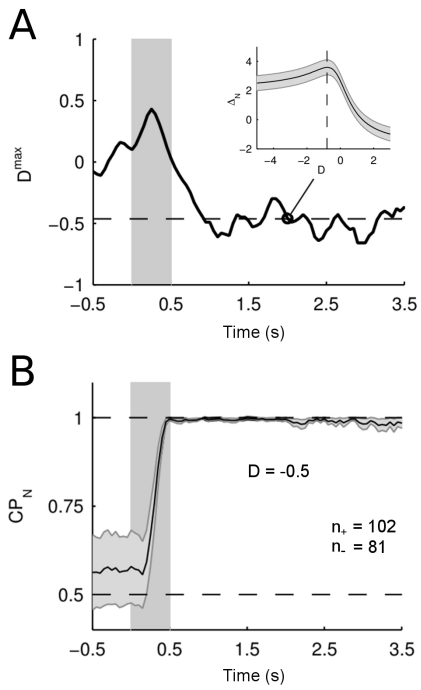

Figure 3.5: **Delay-activity neurons predict unambiguously the behavioral report.(A)** Temporal profile of values of D that maximize the CP_N in each 250 ms time window independently. Inset shows the location of the maximum at a 250 ms temporal bin centered at $t = 2.0s$ as an example. After a transient regime, the optimal value of D reaches the stationary value $D = -0.5$. **(B)** The CP_N index for $D = -0.5$ is compatible with its maximum possible value of 1 during the entire delay period (see also Figure A.4). Shaded area represents SEM.

hibits during the delay period. Figure 3.5 (bottom) shows the temporal evolution of the CP_N index when the two pools are combined using the stationary value $D = -0.5$. The value $CP_N = 1$ indicates that the population of delay-activity neurons unambiguously predicts the behavioral report during the whole delay period. Notice that because of the difficulties in measuring the entire covariance matrix, we cannot obtain an optimal population variable individually weighting each neuron's firing rate (Equation A.38). However, under the assumption of discrete pools, the CP_N index already reaches its maximum possible value, so the conclusion that this population perfectly predicts the animal's behavior still is valid. We would like to note that the application of our analytical tools does not require any assumption about the role of each pool in the decision process. Even so, a plausible interpretation of the optimal rate combination found for the population of delay-activity neurons is that the activity of the negative pool of PM neurons represents the default decision that the stimulus is absent, because when the stimulus is applied, the activity of these neurons diminishes while the activity of neurons in the positive pool increases.

What factor determines the saturation of the CP_N index? We have noticed before (Equation 3.6) that this index is affected by the choice-conditioned correlation coefficients. In our data, $(N-1)-bar\rho^w >> 1$ so CP_N becomes independent of the number of neurons and the amplification of a single neuron's covariance with behavior is controlled by the inverse of $(1+D^2)\,\bar{\rho}^w + 2D\bar{\rho}^b$ (the dependence of CP_N on the number of neurons is discussed in the Appendix A, Figure A.4). We have just seen that for delay-activity neurons $D = -0.5$. In addition, since $\bar{\rho}^b$ is much smaller than $\bar{\rho}^w$ (Figure 3.3B), this expression is mainly determined by $\bar{\rho}^w$, the choice-conditioned correlation coefficient of neurons in the same pool. The smaller $\bar{\rho}^w$, the greater the amplification with respect to the single neuron's CP index.

In view of this result, one may wonder if the population of delay-activity neurons can predict the correct choice before the onset of the tactile stimulus, when the presence of such stimulus is indicated by a separate cue at the beginning of the trial. To answer this question, we decoded the animal's choice from the population of neurons with delay activity in a variation of the task in which the correct response button was illuminated at the beginning of the trial (see Methods). In this control task, monkeys were not required to attend the vibratory stimuli but just to press the cued button at the end of the trial. We hypothesized that if the correct choice is indicated by the light cue at the beginning of the trial and the same neurons are engaged in this variant of the task, the animal's choice could be decoded from the activity of the neural pools even before the application of the stimulus. To test this hypothesis, we evaluated a CP_N index from the population firing rate during stimulus-present and stimulus-absent control trials. We performed this analysis for delay-activity neurons because they are the only population showing significant covariation with behavior during the delay period of the task. Indeed, before stimulus onset, this index is significantly larger than the CP_N in the detection task (Figure 3.6). The large value and stationary profile of the CP_N index in control trials indicates that the choice is made during the prestimulation period. After that, the choice is kept in memory in the form of sustained activity until the end of the delay period.

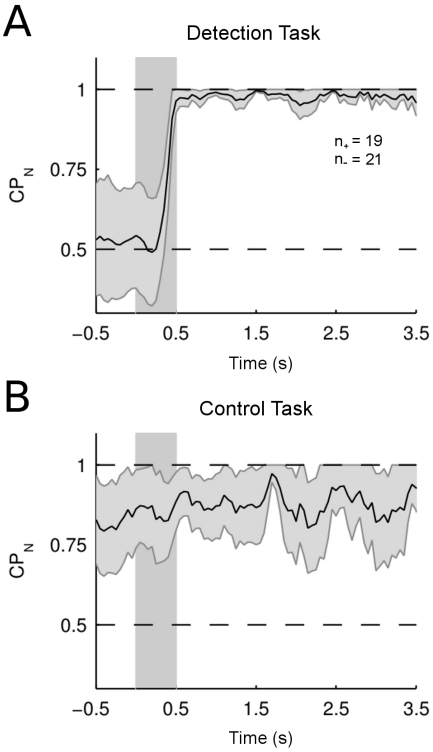

Figure 3.6: **PM neurons reflect the behavioral choice throughout the trial when the correct response is indicated at the start. (A)** Temporal profile of CP_N during the detection task, obtained from a set of delay-activity neurons that were also recorded during cued trials. **(B)** Temporal profile of the CP_N index during control trials. In this case, CP_N was computed using stimulus-present and stimulus-absent correct trials. The large value of CP_N before the stimulus presentation indicates that, in contrast to detection-task trials, here the choice was made during the prestimulation period. Shaded area represents SEM.

3.3 Discussion

In decision-making tasks, the subject's choice results from the coordinated activity of neurons distributed in a large network comprising numerous brain areas. Hence, the decision should be decoded from population variables, based on the spiking activities of the neuronal populations involved. It is then expected that correlated variability between the firing activities of those neurons play a key role in determining the decision. Indeed, the fact that single neurons covary with the subject's report is usually explained by the existence of pairwise correlations between neurons in the neuronal population (Shadlen et al. [1996]).

Experimental and computational studies have shown that recurrent cortical networks can fire in an uncorrelated fashion, even if they receive significant common inputs (Ecker et al. [2010]; Renart et al. [2010]). There is no conflict between the ability of cortical networks to decorrelate the responses of pairs of neurons

and the observation of significant correlated variability in decision-making tasks. Good performance in these tasks requires nonzero correlations evaluated using trials of both choices, but choice-conditioned correlations are not constrained to be high. To gain further insight into these issues, here we estimated the covariance of firing activity with behavior in terms of pairwise noise correlations and found that choice probability depends in a remarkably simple way on these quantities: it is essentially given by the overall noise correlation coefficient (computed from all trial types together) minus the average choice-conditioned correlation. It is the first term (of Equation 3.8) that is needed to explain why single neurons have a significant CP, while it is the second term that, according to the decorrelating effect of recurrent cortical networks, has to be small. In fact, the negative contribution of this term shows that nonzero choice-conditioned noise correlations always decrease the covariance between firing activity and behavior. The theoretical understanding of this issue is verified with great accuracy by the analysis of experimental data. CP is indeed explained by the difference between the two correlation types (Figure 3.3C). Mean choice-conditioned correlations are modulated during the time course of the task, with values in the range between 0.2 and 0.05 (Figure 3.3B). At least part of these correlations can be explained by the existence of an internally generated signal fluctuating from trial to trial (Carnevale et al. [2012]). The smallest observed value could still contain this effect.

So, our study shows that correlations need to be considered if one is to analyze covariations of population firing rate variables with the subject's report. In perceptual decision-making tasks, the perceptual report results from neural processes distributed over several interacting neuronal populations and over a number of brain regions (Hernández et al. [2010]; Siegel et al. [2011]). The relevant measure of covariance between firing activity and behavior is a generalized CP index defined as a combination of firing rates from appropriate pools of neurons. The hypothesis behind this proposal is that when two or more neuronal populations transiently cooperate in the process of forming the decision, they produce a combined signal that accurately predicts the behavioral report. Hence, the most relevant mixture of the populations firing rates can be obtained by

maximizing the covariance between the linear combination of rates and behavior. Following these ideas, we found that populations of PM neurons active during the delay period of a detection task unequivocally predict the animal's decision (Figure 3.4E). When the animal is cued at the beginning of the trial about the correct choice, full covariance with behavior may be reached even before a (now irrelevant) stimulus is presented (Figure 3.6B).

In summary, we developed tools to evaluate the CP index from the correlation structure of a network without assuming any decision rule. The tools can be applied to both sensory and frontal areas, as we showed in the analysis of S2 and PM neurons. We then generalized the use of choice probability indices to population variables and found a way to evaluate them based on data from simultaneously recorded pairs of neurons. This allowed us to propose a procedure for determining how the activity of neurons in different populations should be combined to optimally predict the subject's behavior (based on the linear Fishers discriminant). Finally, we were able to find a population variable that fully covaries with behavior. These analytical tools may be employed to study the dynamics of cortical networks engaged in keeping relevant information in short-term memory. In the detection task, our results suggest that by the end of the stimulation period, the decision is already made and it is maintained in short-term memory during the entire delay period. In a somatosensory discrimination task with two delay periods (Brody et al. [2003]; Hernández et al. [2010]; Lemus et al. [2007, 2009]; Romo et al. [1999]), a first vibratory stimulus is kept in memory during the first of these intervals and this memory is later compared with a second stimulus. After a second delay period, the subject has to report which of the two stimuli vibrated with the highest frequency. It was found that during this interval, populations of PM neurons maintain in memory the frequency of the two stimuli, even if the decision could already have been made, as in the detection task (Hernández et al. [2010]; Lemus et al. [2007]). Does this mean that the decision continues to be elaborated, based on the information about the two stimuli maintained in working memory? Analyzing how CP_N evolves in time could answer these dynamical issues.

Neurons recorded in the detection task could be assigned to one of a small set

of pools, a property that simplified the study of the CP_N index. In discrimination tasks, however, neuronal firing rates are rather heterogeneous (Jun et al. [2010]) and the use of general equations where neurons contribute with different weights could be needed Equations A.38 and A.39). In addition to determining whether populations of PM neurons stably maintain a decision after presentation of the second stimulus, estimating the weights with which neurons contribute to the decision would provide a method to rank neurons according to their relevance in the task. This issue could be studied using data sets in which a few tens of neurons have been recorded simultaneously (Hernández et al. [2008]). The ideas developed in this work could be applied to study brain functions other than detection of sensory stimuli. Cohen and Maunsell [2011] have noticed that attentional fluctuations are associated with fluctuations in psychophysical performance. To reach this conclusion, these authors evaluated an ROC index based on a population firing rate variable defined in terms of two attentional states, a quantity somewhat similar to our CP_N index. Noise correlations are reduced by both spatial and feature attention, a fact that is assumed to have a positive effect on stimulus coding (Cohen and Maunsell [2010, 2011]). Our results show that noise correlations are also relevant to explain the covariance between neuronal activity and choice and that small choice-conditioned correlated variability is needed to achieve a larger covariance with behavior both for single cells and for neuronal populations (Figures 3.3C and 3.4). Whether this is also true for neurons in higher visual areas such as V4 would require an analysis of the time course of correlations.

The simplicity of the approach presented here makes it feasible to study a wide spectrum of problems. From a theoretical viewpoint, we provided an intuitive framework to understand how first- and second-order statistics affect the relationship between network firing activity and behavior, which can be used to further develop computational methods. From a data analysis perspective, our approach could help to reveal how several cortical areas contribute and collaborate in the decision-making process.

3.4 Methods

3.4.1 Detection Task

Data for this analysis were obtained from two earlier studies (de Lafuente and Romo [2005, 2006]). Stimuli were delivered to the skin of the distal segment of one digit of the restrained hand, via a computer-controlled stimulator (BME Systems; 2 mm round tip). Initial probe indentation was 500 mm. Vibrotactil stimuli consisted of trains of 20 Hz mechanical sinusoids with amplitudes of $2.3 - 34.6mm$. These were interleaved with an equal number of trials where no mechanical vibrations were delivered to the skin (amplitude = 0). Animals pressed one of two buttons to indicate stimulus present (left button) or stimulus absent (right button). They were rewarded with a drop of liquid for correct responses.

3.4.2 Recordings

The activity of pairs of neurons were simultaneously recorded from the same cortical area including secondary somatosensory cortex (S2), ventral premotor cortex (VPc) on the left hemisphere, and dorsal premotor cortex (DPc) and medial premotor cortex (MPc), bilaterally. Pairs from premotor cortices were not distinguished in this report. Trials in the control light task proceeded exactly as described in Figure 3.1A, except that at the probe down, the correct target button was illuminated. Vibrotactile stimuli were delivered while the light was kept on; then, at the probe up, the light was turned off. The monkey was rewarded for pressing the previously illuminated button. Detailed description of the experimental techniques was described in de Lafuente and Romo [2005, 2006]. Animals were handled in accordance with the standards of the NIH and the Society for Neuroscience.

3.4.3 Data analysis

Statistical properties of the firing activity

Statistical properties of the firing activity (firing rate, firing rate variance, and correlation coefficient) were computed for each neuron or pair of neurons as a function of time using 250 ms sliding window displaced every 50 ms. Trials were aligned to the time of stimulus onset. Firing rate, r(t), was calculated as the number of spikes in one sliding window divided by the its temporal length. Mean firing rate in condition c, mc(t), was obtained averaging over all trials of this condition. The SE of the mean firing rate was computed as the SD over trials divided by the square root of the number of trials. Variance of the firing rate in trials of condition c, sc(t), was obtained using

$$\sigma_c^2(t) = \langle r_i(t)^2 \rangle_c - \langle r_i(t) \rangle_c^2 \tag{3.10}$$

where c indicates average over trials of condition c. The SE of the variance was $\sigma_c^2(t)/\sqrt{2/\left(N_c - 1\right)}$, where N_c is the number of trials of condition c.

Correlation coefficients of the firing rates of a pair of neurons (i,j), in trials of condition c, were calculated following,

$$\rho^c(t) = \frac{\langle r_i(t) r_j(t) \rangle_c - \langle r_i(t) \rangle_c \langle r_j(t) \rangle_c}{\sqrt{\sigma_{i,c}^2(t) \sigma_{j,c}^2(t)}} \tag{3.11}$$

Statistical properties of the firing activity were computed only from neural recordings with at least five trials.

Measures of covariance with behavior

Choice probability indices were calculated using Equation 3.1 and computed by direct numerical evaluation (Figure 3.2). The Complementary error function

(erfc) in Equation 3.1 was computed numerically (MATLAB, MathWorks). The SE of the CP calculated using Equation 1 was obtained by propagation of the SEs of the firing rates and firing rate variances over the formula. Direct numerical evaluation of choice probability was obtained using methods of signal detection theory (Green et al. [1966]) implemented with custom software written using MATLAB (MathWorks). Population-averaged $CP_{2,w}$ index in Figure 3.2 was computed evaluating Equation A.16 for each pair of neurons and averaging over all pairs of neurons within the same neural pool. Similarly, population-averaged $CP_{2,b}$ index in Figures 3.4B and 3.4C was obtained for each neuronal pair of neurons belonging to different neural pools by evaluating Equation A.22 and averaging over the corresponding population of pairs.

CP_N in Figures 3.4D and 3.4E and in Figures 3.5 and 3.6 was computed from population-averaged statistical properties of the firing activity using Equation A.31. Population-averaged quantities were estimated pooling neurons and neuron pairs across different recording sessions. Both $CP_{2,b}$ and CP_N were computed for different linear combinations of pool rates and the optimal one was defined as that with maximum value of $D_{2,b}$ or D_N, respectively. This gives the optimal value of the coefficient D.

Full noise correlations and choice-conditioned noise correlations

In Figure 3.3B, the correlation coefficients of the firing rates computed with all trials (R) were obtained numerically using Equation 11 and analytically by evaluating Equation A.48. Both were computed for each pair of delay-activity PM neurons and then averaged over the population of pairs.

Population-averaged choice probability index as a function of correlation coefficients (Figure 3.3C) was computed evaluating Equations 3.8 for each pair within the population of delay-activity PM neurons and then averaged over the population of pairs.

Chapter 4

Dynamic control of response criterion to incorporate temporal expectations

4.1 Introduction

One of the main challenges of cognitive neuroscience is to understand how external sensory stimuli and internal brain states interact to give rise to perception (Romo and de Lafuente [2013]). Internal states are believed to reflect acquired experience that can be used for making the best sense of our sensory inputs (Gilbert and Sigman [2007]). During perceptual decisions, for example, the brain uses previous knowledge to transform the noisy sensory evidence into the percepts on which decisions are based (Forstmann et al. [2010]; Hanks et al. [2011]; Rao et al. [2012]; Ratcliff and McKoon [2008]; Simen et al. [2009]; Summerfield and Koechlin [2008]). In this study, we explore the dynamic nature of these internal states by asking how previous information about the timing of sensory evidence is incorporated in the decision-making process. We combine computational modeling with neurophysiological and behavioral data recorded while monkeys performed a somatosensory detection task (de Lafuente and Romo [2005, 2006]).

Subjects performing a decision-making task can benefit from the use of temporal expectations (Coull and Nobre [2008]) at multiple stages of the sensorimotor transformation (Nobre et al. [2007]): (1) perception can be enhanced by increasing sensory accuracy at the relevant times (Correa et al. [2005]; Ghose and Bearl [2010]; Ghose and Maunsell [2002]; Jaramillo and Zador [2011]; Rohenkohl et al. [2012]); (2) the response criterion - the subject's internal rule to decide whether or not to report a stimulus - can be modulated to incorporate prior information without changes in the sensory representation (Katzner et al. [2012]); (3) Motor readiness can be heightened, increasing the response speed in reaction-time tasks (Nobre [2001]; Scheibe et al. [2009]). These studies have found neurophysiological evidence for the use of temporal information in the sensory and motor stages. However, little is known about the neural mechanisms that underlie the use of timing at intermediate stages of the sensorimotor transformation.

We address this intermediate step by analyzing recordings of premotor cortex neurons from monkeys performing a detection task with variable stimulus onset times (Figure 4.1a. See de Lafuente and Romo [2005, 2006]). The task's temporal structure dictated that the stimulus only arrived within a 2 s temporal window but not before or after (Figure 4.1b). We asked whether monkeys can infer and take advantage of this temporal structure to increase performance. One possible way to incorporate this knowledge is to modulate the response criterion (the amount of sensory evidence required to produce a stimulus-present response) over the time course of the trial (Figure 4.1c). An efficient modulation of the criterion is to raise it outside the possible stimulation window to avoid false positive outcomes, and lower it within the window to allow correct detections. The exact shape of the response criterion within the possible stimulation window depends on the animal's inference about the underlying distribution of stimulus onset times (the subjective hazard function, see Janssen and Shadlen [2005]; Luce [1986] and Discussion).

How could a population of neurons implement such a mechanism? If we consider the abstract high dimensional space of neural activity, the threshold to commit to a decision can be pictured as a boundary, that once crossed, triggers perceptual detection (Figure 4.1d, the dynamical systems term for the boundary is

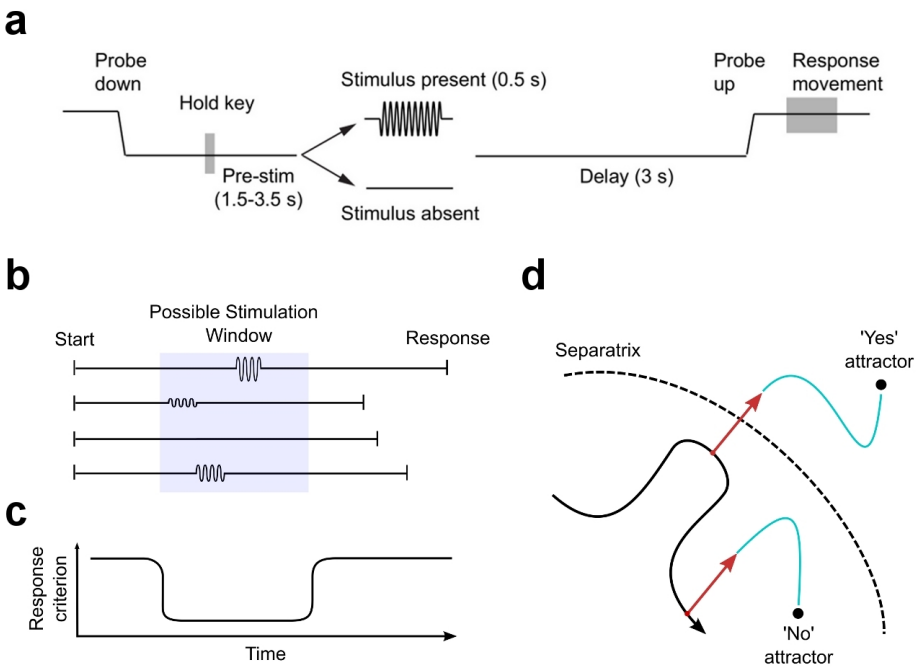

Figure 4.1: **Detection task and dynamical response criterion (a)** Behavioral task represented by the vertical position of the mechanical probe during a trial. The stimulator indented the skin of one fingertip of the restrained hand ('probe down') and the monkey reacted by placing its free hand on an immovable key ('hold key'). After a variable prestimulus period (1.5-3.5 s), a vibratory 0.5 s stimulus was presented on half of the trials. At the end of a fixed delay period (3 s), the stimulator moved up ('probe up'), instructing the monkey to make a response movement to one of two push buttons. The pressed button indicated whether or not the monkey felt the stimulus. **(b)** The variability in stimulus onset times and the fixed delay period defined a 2 s temporal window of possible stimulation. No stimulus was delivered before 1.5 s or after 3.5 s from the 'hold key' event. The window of possible stimulation was not explicitly cued to the animal. **(c)** A possible mechanism to efficiently solve the task requires modulating the response criterion (the strength of sensory evidence required to produce a stimulus-present response) over time. Outside the possible stimulation window, the response criterion is high to avoid false positives. Within the window, the response criterion decreases to allow correct detections. **(d)** The mechanism described in (c) could be dynamically implemented by a separatrix in the neural space, dividing the basins of attraction of two attractors. The black trace is a trajectory of a correct rejection trial. The blue traces represent a Hit (ending in the 'yes' attractor) or a Miss trial (ending in the 'no' attractor). The distance from the current neural state to the separatrix at each point in time represents the response criterion.

a separatrix). The response criterion is simply the distance from the current state of the network to that boundary. Temporal expectations can then be manifested via the trajectory of neural dynamics while the monkey is waiting for a stimulus - drawing closer to the boundary when the stimulus is expected and vice versa (Figure 4.1d).

In this work, we present experimental and modeling evidence in favor of this dynamical mechanism. We start by using the timing of false alarms to infer the dynamics of the response criterion. To obtain these times, we develop a novel method to detect neural correlates of false alarms and find that indeed their probability increases during the period of possible stimulation. Following the intuition outlined above, we analyze the dynamics of correct rejection trials - as these encapsulate the 'waiting for a stimulus' condition. We show that in these trials the neural trajectory is modulated precisely during the period of possible stimulation. Finally, we derive a model by training a recurrent network to perform an analogous detection task. We find that the model is able to infer the task's temporal structure and using it, we unveil the explicit dynamical implementation of the proposed neural mechanism.

4.2 Results

Monkeys were trained to detect a weak mechanical vibration of variable amplitude applied to one of their fingertips. Reward was provided for correctly reporting the presence (Hit trials) or absence (correct rejection trials, CR) of the stimulus. In contrast, no reward was delivered during incorrect trials, which arose either from missing a stimulus (Miss trials) or reporting a false positive (false alarm trials, FA). The stimulus onset time varied from trial to trial between 1.5 s and 3.5 s after the 'hold key' event (Figure 4.1a). Following stimulation (or absence of) monkeys had to wait for a 3 s delay period until a cue indicated to report their decision. Because of this temporal structure, we expect subjects to modulate their response criterion to benefit from the fact that no stimulus arrived before 1.5 s or after 3.5 s (Figure 4.1b,c).

A modulation in the response criterion has predictable consequences on behavior. A higher response criterion leads to an increase in the frequency of stimulus-absent responses while a lower response criterion implies an increase in the frequency of stimulus-present ones. Therefore, evidence of change in response criterion over time could be obtained by estimating the frequency of stimulus-present responses as a function of time. However, in a delayed-response task there is no behavioral information about the exact time at which the subject reached a decision and, therefore, it is not possible to estimate a time-varying response criterion from behavioral data.

Nonetheless, in any two-alternative forced choice task, a decision represents a commitment to one of the two possible alternatives. Thus, we hypothesized that information about the timing of the subject's decision could be found in the neural activity. Premotor cortex (PMc) activity was previously shown to correlate more with the subject's perceptual decision than with the physical properties of the stimulus (de Lafuente and Romo [2005]). In fact, in Chapter 3, I showed that the subject's decision can be unambiguously decoded from a population of neurons in this cortical area (Carnevale et al. [2013]). Moreover, premotor cortex activity was previously shown to reflect an internal component of the decision process (Carnevale et al. [2012]). Therefore, we set to find information about the subject's response criterion from the firing activity of PMc neurons. We analyzed an experimental data set from two earlier studies (de Lafuente and Romo [2005, 2006]) which included single-neurons and small sets of simultaneous neuronal recordings (up to 6 cells with a median of 2), summing to a total of 384 extracellularly recorded neurons (see Methods).

4.2.1 False alarms as a window onto the response criterion's dynamic

As stated above, if monkeys modulate their response criterion during the time course of a trial, this should be reflected in the probability of producing a false alarm over time. Here we set to find this information from single trial neural activity. We assume that the decision process carried out in every trial led to a

commitment to one of the two possible alternatives. Therefore, we expect some neurons to present irreversible and stereotypical activity profiles reflecting this commitment. Indeed, we noticed that premotor cortex neurons during single FA trials show temporally localized fluctuations in their firing rate (Figure 4.2a, left). The profile of these events resembles the neuron's responses evoked by the vibrotactile sensory stimulation (Figure 4.2a, right). Furthermore, they occur at the same time in simultaneously recorded neurons, suggesting that they correspond to the same perceptual event. Taken together, this suggests that these fluctuations, which we called FA events, are neural correlates of false alarms.

We devised a method to detect the time of production of FA events from single FA trials (Figure 4.2b, see Methods). For each neuron, we used the average firing rate in Hit trials to define a 1 s template representing the neuron's specific response to the external stimulation (Figure 4.2b, inset). Applying this template to each individual FA trial, we searched for similar firing profiles, providing putative FA event times (Figure 4.2b). By realigning the FA trials according to the detected times, we obtained an average response resembling that of the Hit trials (Figure 4.2c, blue and green traces). Importantly, this is true not only during the 1 s period used as a template, but also during the remaining 2 s of the delay period, consistent with the idea that what we detected is a stereotypical activity profile equivalent to the one evoked by the external stimulus. Moreover, the this applies to neurons with very diverse firing temporal profiles (see Figure B.1).

Due to the noisy nature of single trial data, our template-matching algorithm produces a large amount of false detections. In particular, simply observing the average of the realigned FA trials (Figure 4.2c) suffers from a circular logic - we may pick out events from noise and by definition get a similar waveform after averaging the realigned trials. To validate the significance of the detected events, we used the activity of simultaneously recorded neurons. If the FA events are neural correlates of FAs, they should occur at the same time on different neurons. For each trial, we compared the FA event times obtained independently from two different neurons. The histogram in Figure 4.2d (green bars) shows the frequency of FA event's time differences over the entire set of FA trials. A significant fraction of FA events were detected at the same time compared to

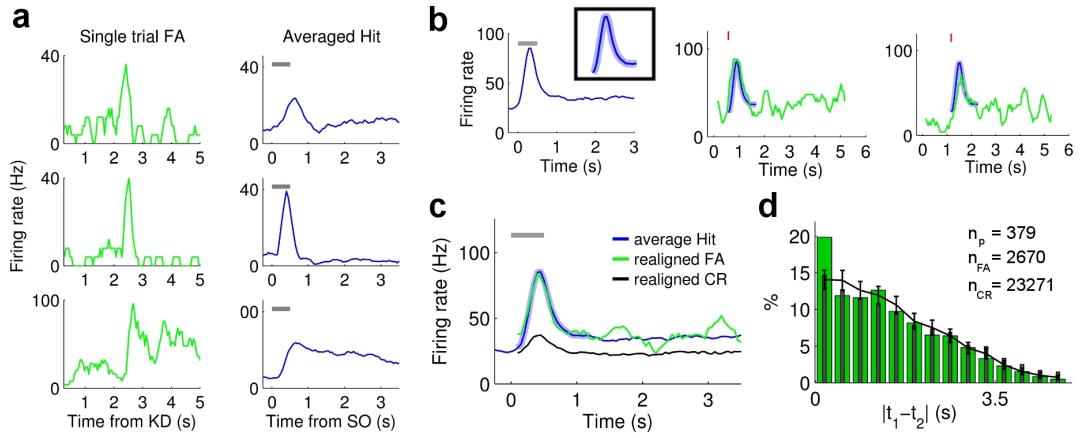

Figure 4.2: **Detection of False Alarm events by template-matching (a)** Firing rate of three simultaneously recorded neurons during a single FA trial (green traces) and average response of the same neurons during Hit trials (blue traces). The shaded bar indicates the stimulation period. **(b)** We used a 1 s segment of the averaged activity during Hit trials (left, blue trace) as a template (inset) to detect FA events in single FA trials. FA events were identified in single FA trials (middle and right, green traces) on the basis of the mean squared error between the single FA trial firing rate and the template. Red lines indicate the start of the template. **(c)** The average activity over FA trials realigned according to the times of detected events (green trace) matches the average over Hit trials (blue trace) even outside of the period used as template. In contrast, the same method applied to CR trials produces a much weaker match.**(d)** Histogram of differences in the detected FA times from pairs of simultaneously recorded neurons. A significant fraction of FA trials was detected at the same time compared to CR trials (black bars) and chance level chance level (black line, $p<0.001$). Chance level was obtained by shuffling the trials to disrupt the correspondence between detected FA events in simultaneously recorded neurons. Error bars indicate 95% confidence intervals. n_p is the number of neural pairs, n_{FA} and n_{CR} are the number of FA and CR trials respectively.

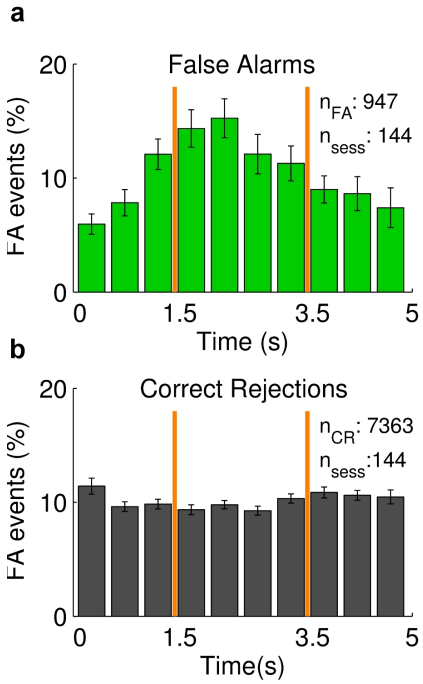

Figure 4.3: **Probability of false alarm over time(a)** Mean relative frequency of detected FA events over time during the time course of the trial. The probability of a FA event increases during the period of possible stimulation (within orange lines). Relative frequency was calculated as the portion of FA events detected in each time bin relative to the number of FA trials in which a FA event was detected at any time bin. The mean histogram was obtained by averaging across $n = 947$ FA trials distributed in $n = 144$ sessions. Error bars represent s.e.m. **(b)** Same panel **(a)** for CR trials.

chance level (compare first green time bin to black line). We applied the same template-matching algorithm to CR trials to further control for circular logic, as they presumably have the same noisy nature as FA trials but do not present FA events. CR trials revealed both a weaker agreement with the average Hit firing rate (Figure 4.2c, black trace), as well as no significant number of simultaneous events (Figure 4.2d, black bars). Taken together, these results suggest that at least a subset of FA trials can be explained by an event that is localized in time and that triggers an irreversible and stereotypical neural activity pattern, equivalent to the one evoked by external stimulation.

Under the hypothesis that these FA events generate a false percept, the estimation of the times at which these events are produced allows to obtain the probability of a false alarm over the time course of the trial. By using those trials in which FA events were detected in two or more neurons simultaneously (first bin of Figure 4.2d), we computed the frequency of detected events across time. The resulting probability is not uniform (Figure 4.3a, green bars) but reaches a maximum during the period of possible stimulation (Figure 4.3a, orange lines).

In contrast, the same quantity obtained for CR trials, used as a control, revealed no modulation during this period (Figure 4.3b, black bars). The increase in the probability of FA during the possible stimulation window is consistent with a decrease in the subject's response criterion when the stimulus is more likely to arrive. Figure 4.3a suggests that monkeys are able to infer the task's temporal structure and make use of this knowledge to modulate their response criterion according to the stimulation probability.

4.2.2 Premotor cortex dynamics suggests a neural mechanism for modulating the response criterion

So far we showed that monkeys used previously acquired temporal information when performing the vibrotactile detection task. In the previous section we used FA trials to show that the subject's response criterion was modulated over the time course of the trial. In this section we aim to find signatures of the dynamics of this modulation in the neural activity. We analyze CR trials to show that the activity in this condition reflects the subject's expectations about the probability of stimulation over time.

The activity of neurons in PMc was previously shown to reflect the subject's perceptual judgment about the presence or absence of the stimulus during the vibrotactile detection task (de Lafuente and Romo [2005]). Several pieces of evidence were examined in that study. First, PMc neurons responded in an all-or-none manner, only weakly modulated by the amplitude of the stimulus. Second, when presented with a fixed near-threshold stimulus, PMc activity strongly correlated with the subject's choice. Third, reversing the direction of the arm movements in control experiments did not change the activity of PMc. Fourth, when PMc was electrically micro-stimulated, the probability of stimulus-present responses was higher than when only the mechanical stimuli was presented. In fact, in Chapter 3, I showed that the subject's decision could be unambiguously predicted from the activity of populations of PMc neurons (Carnevale et al. [2013]).

If PMc activity represents the subject's perceptual judgments, we expect that a modulation in the subject's response criterion will be reflected in the firing rate of PMc neurons. However, the response of PMc is quite heterogeneous across neurons (de Lafuente and Romo [2006]). When presented with a supra-threshold stimulus, some neurons increased their firing rate while others tended to decrease it. Moreover, the temporal profile of PMc neural responses was also diverse. Some neurons responded only during stimulus presentation while others showed persistent activity or even ramping profiles during the delay period of the task (see also Figure 4.3a). In face of this heterogeneity, it is not trivial to predict how a modulated response criterion would be reflected in each single neuron's activity.

State-space analysis was shown to be a useful tool to study the neural dynamics at the population level (Mante et al. [2013]; Shenoy et al. [2013]; Stokes et al. [2013]). In this framework, the activity of a population of N neurons at each point in time is represented as an N-dimensional point in the space spanned by each neuron's activity. The population activity across time defines a trajectory within this space. The set of neural trajectories often occupies a low-dimensional subspace within the space of possible activities, and various methods can be used to visualize it. We generated 132-dimensional trajectories by combining neural data mostly recorded separately (the number of neurons was limited by the need to match conditions between different recording sessions, see Methods). Then, we projected these trajectories onto two task related axes - tuning to stimulus amplitude and tuning to a detection event. The former was defined by regressing each neuron's trial-to-trial response to the stimulus amplitude. The latter was defined as the vector connecting the network state just before the application of the stimulus to the network state at the end of the delay period during hit trials.

Figure 4.4 shows the average neural trajectories during Hit, Miss and CR trials projected onto these axes. We omit the FA condition because, as we showed before, the neural activity during each individual FA trial is equivalent to the one during a Hit and thus the averaged FA trajectory is actually composed by many individual FA events misaligned and distributed over time. For the Hit condition, we discarded weak amplitude trials to avoid the inclusion of possible FA events,

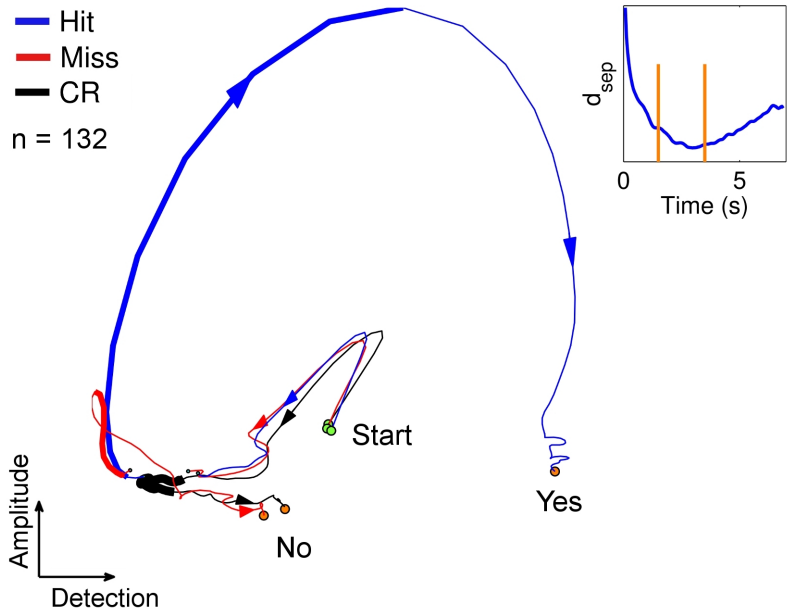

Figure 4.4: **Two-dimensional projection of the population dynamics** Average neural trajectories during Hit (blue), Miss (red) and CR (black) trials projected onto two task-related axes (stimulus amplitude and stimulus detection). The trajectories are plotted from the beginning of the trial (green circles) to end of the delay period (orange circles). Stimulus-present conditions are plotted until 1.5 s and realigned at the stimulus onset time. Thick blue and red traces indicate the period of stimulation. The thick black line denotes the possible stimulation window (1.5 s to 3.5s). Units are arbitrary. The inset is the N-dimensional euclidean distance between the CR neural trajectory and the neural state at the stimulus offset time during the Miss condition (end of the thick red trace), as an estimate of the distance to the separatrix over time. See Figure B.2 for the same analysis performed for each subject separately.

while for the Miss condition, we dismissed strong amplitude trials to avoid errors due to lapsus. We plotted each trajectory from the beginning of the trial to the end of the delay period (Figure 4.4, green and orange circles, respectively). Due to the variable stimulus onset times, stimulus present conditions (Hit and Miss) are aligned first to the 'hold key' and then to the 'stimulus onset' times.

As expected, the neural trajectories for the three conditions overlap at the beginning of the trials. In Hit and Miss conditions (blue and red traces) the application of the stimulus drives the network in the stimulus direction. The stimulus on Hit trials moves the network further than in the Miss condition. Afterwards, during the delay period, the network evolves into two different trajectories. The endpoints of these trajectories (orange circles) represent the final states of the network, presumably corresponding to the stimulus-present and stimulus-absent choices (compare to Figure 4.1d).

We used the same axes to project the population activity during CR trials. In this condition the monkey waited for a stimulus that never came, and then correctly reported its absence. Thus, the monkey's knowledge about the task's temporal structure and the resulting expectations about the stimulus arrival should be reflected in this condition. The neural trajectory during CR trials is shown in Figure 4.4 (black trace). Interestingly, it presents a strong modulation precisely during the period of possible stimulation (thick black line). After 3.5s from the beginning of the trial, the network state evolves to the same final state as the trajectory in the Miss condition (stimulus-absent state). Importantly, the projection axes were not selected 'ad hoc' to see this modulation. In fact, a similar modulation is observed when the neural trajectories are plotted onto the two principal components of the data. (see Figure B.3).

Figure 4.4 is consistent with our proposed dynamical mechanism (Figure 4.1d). The modulation observed in CR trials during the possible stimulation window can be a signature of the network approaching a separatrix beyond which the dynamics leads to a stimulus-present response. During the possible stimulation window, this distance should decrease, lowering the response criterion when the stimulus is more likely to come and then it should increase again. While the location of such a separatrix cannot be obtained from the recorded neural activity,

the network state just after the offset of the stimulus during a miss trial should be below and close to it. We can then estimate the distance to the separatrix as the euclidean distance in the high-dimensional space between this state (the average neural activity during miss trials at stimulus offset) and the neural trajectory during CR trials. This measure indeed decreases during the period of possible stimulation (Figure 4.4, inset).

Summarizing, the state of the network while the monkey is waiting for a stimulus is not stationary. In contrast, Figure 4.4 suggests that the neural trajectory could intrinsically encode the temporal information about the probability of stimulation over time. While the subject is waiting for the stimulus, the neural trajectory is determined by the neural dynamics. Therefore, temporal expectations that the subject built during training, might be stored in the internal dynamics of the neural population.

4.2.3 A recurrent network unveils the dynamical implementation of response criterion modulation

What dynamical mechanism supports the use of prior temporal information during perceptual detection? We used a recurrent neural network model to answer this question. Starting with a random recurrent network, we trained it to perform a simplified version of the experimental task. After verifying that the model is able to solve the task, we analyzed the solution achieved. We were specially interested in whether the developed solution makes use of temporal information. We asked if the network is able to benefit from temporal information acquired during the training phase and, if so, what are the dynamical mechanisms by which this information is integrated with the sensory evidence to detect the presence of a stimulus.

Our model is a recurrent neural network of rate units, provided with two inputs and one output (Figure 4.5a, see Methods). The first input is used to signal the start of a new trial, while the second one represents the sensory channel via which the stimulus is applied. The stimulus is modeled as a pulse proportional

to the vibration's amplitude, embedded in a noisy background. In each trial the decision about the presence or absence of the stimulus is indicated by the value of the output during a readout interval. Trials begin with a start cue applied through one of the input channels. After a variable pre-stimulus time, on half of the trials, a stimulus is presented through the sensory channel. Then, after a fixed delay period, the decision is extracted from the network's output through a linear read out. Trials are simulated in a continuous manner, without any reset between them– the start cue input provides the time reference within a trial.

After randomly initializing all synaptic weights, we trained the recurrent network to solve the task. During the training phase we used the FORCE algorithm to change the output connections (Sussillo and Abbott [2009]). Although these are the only plastic weights, because of the feedback loop this change results in a rank-one perturbation to the effective recurrent weight matrix, therefore changing the dynamics of the network (Sussillo and Abbott [2012], see Methods). Training is controlled by a teaching signal representing the desired output in each trial. Since we want to find out whether the strategy developed during training makes use of the timing of the sensory evidence, we provided no explicit information about the probability of stimulation over time. The teaching signal was restricted to the behavioral outcome on each trial (Figure 4.5b) -an analogous information to the one that the monkeys receive in the experimental setup.

The resulting network learns to solve the task. Performance is controlled by the amount of noise in the sensory channel, so, once trained, we calibrated the noise amplitude to approximately reproduce the averaged experimental psychometric function (Figure 4.5b; compare to de Lafuente and Romo [2005], Figure 1c). Then, we asked if the network is able to infer the task's temporal structure and use this information to perform the task. Because this is a model, we are able to test the network's behavior on a large number of stimuli without inducing any learning. Thus, we systematically probed the trained model with variable amplitude stimuli applied at different times from the beginning of the trial. We followed a bisection protocol to find the lowest stimulus amplitude which drives the network to a stimulus-present response (see Methods). This quantity, which represents the model's response criterion, is not fixed but decreases during the

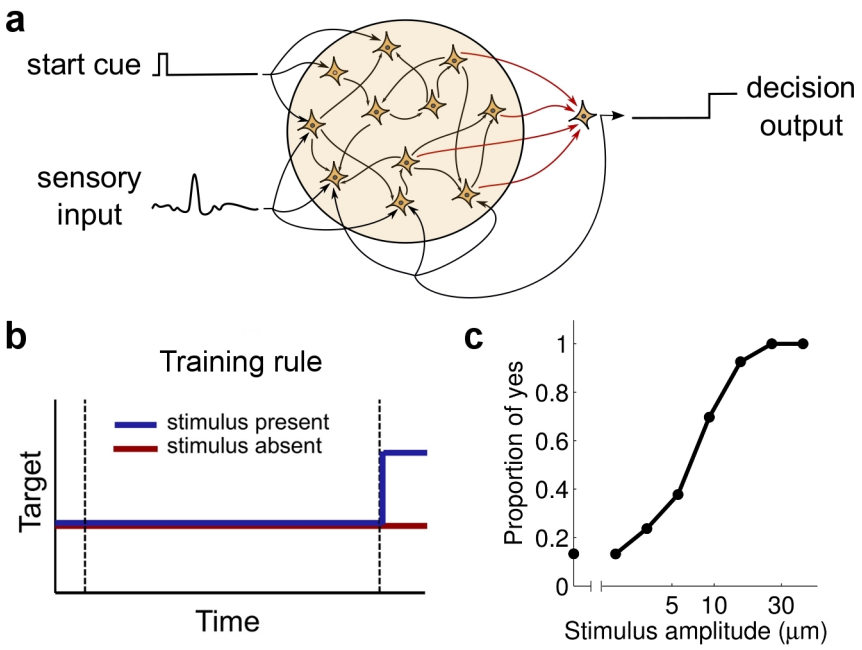

Figure 4.5: **A recurrent network model learns to solve the task (a)** Recurrent network model of rate units provided with a start cue input, a noisy sensory channel and a decision output. The start cue indicates the beginning of a new trial. The stimulus is modeled as a pulse corrupted by noise. The decision is extracted from a linear combination of rates after the delay period. We trained the initially random network by changing the output connections. Because of the feedback loop, this effectively alters the recurrent dynamics of the network.**(b)** Target signal of the FORCE algorithm. The information provided during training was restricted to the behavioral outcome on each trial. Thus, no information about the probability of stimulation over time was given during training. **(c)** 'Psychometric' function of the trained model obtained as the frequency of stimulus-present responses as a function of stimulus amplitude.

period of possible stimulation used during training (Figure 4.6a). To verify the dependence of this measure on the statistics upon which the network was trained, we repeated this procedure for different possible stimulation windows and observed that the response criterion modulated accordingly (Figure 4.6a, inset). This modulation was also revealed when we applied the same template-matching algorithm we used on the experimental data to the FA trials produced by the simulation.The probability of false alarms increases during the possible stimulation window (compare Figures 4.6b and 4.3a).

Finally, we set to understand what dynamical mechanism, developed during training, supports the modulation of the response criterion during the time course of the trial. To do so, we reverse-engineered the network by looking for slow and fixed points, and analyzing the linear dynamics around them (Sussillo and Barak [2013]). We found that the network's dynamics is governed by three fixed points, two stable and one unstable (Figure 4.7). The two attractors correspond to each of the possible decision outcomes ('yes' and 'no' attractors). The third fixed point presents only one unstable eigenvector, and therefore is a saddle point. This saddle point defines a separatrix between the basins of attraction of the two stable fixed points (Figure 4.7, gray dots; see also Figure B.4). At any point in time, strong enough sensory input can make the network cross the separatrix and travel to the stimulus-present attractor, resulting in a hit trial (Figure 4.7, blue trace). In contrast, a weak sensory input will fail to drive the network across the separatrix, resulting in a missed stimulus (Figure 4.7, red trace). The distance between the network state and the separatrix controls how strong the sensory input must be to produce a stimulus-present response (Figure 4.7, inset; compare with Figure 4.6a). Therefore, the network's response criterion can be modulated by controlling the state of the network relative to the separatrix. The neural trajectory that the network developed after the training phase, relative to the separatrix, represents the response criterion at each point in time and incorporates the knowledge acquired about the timing of sensory evidence.

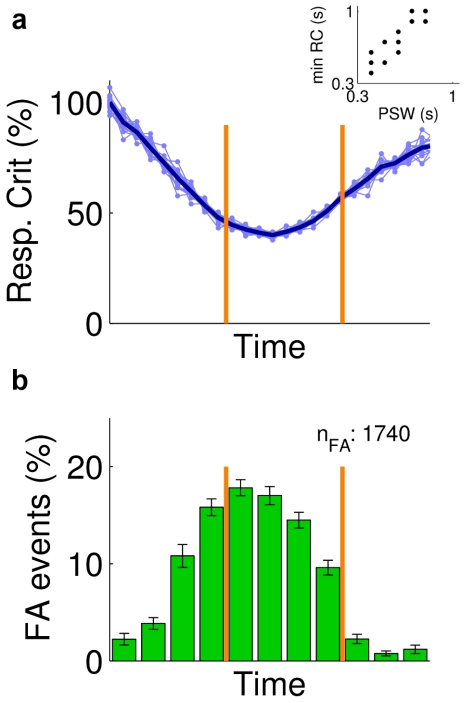

Figure 4.6: **The network infers the window of possible stimulation (a)** The response criterion, defined as the lowest stimulus amplitude that drives the network to a stimulus-present response, decreases during the period of possible stimulation (within orange lines). The response criterion was obtained by systematically probing the network with a bisection protocol at each time to find 'borderline' stimulus amplitudes. Thin lines represent single realizations of this protocol. Thick line is the mean of $n = 10$ realizations. The response criterion was normalized with its maximum value during the trial. Inset shows the results of training networks with different possible stimulation windows. PSW is the center of the possible stimulation window used during training; min RC is the time in which the response criterion reaches its minimum value. **(b)** Mean relative frequency of detected FA events over time in the model obtained by the same template-matching algorithm used for the experimental data. The probability of producing a FA increases during the period of possible stimulation (within orange lines). Relative frequency is defined as in Figure 3. The mean histogram was obtained by averaging across sessions. Error bars represent s.e.m. n_{fa}, number of FA trials.

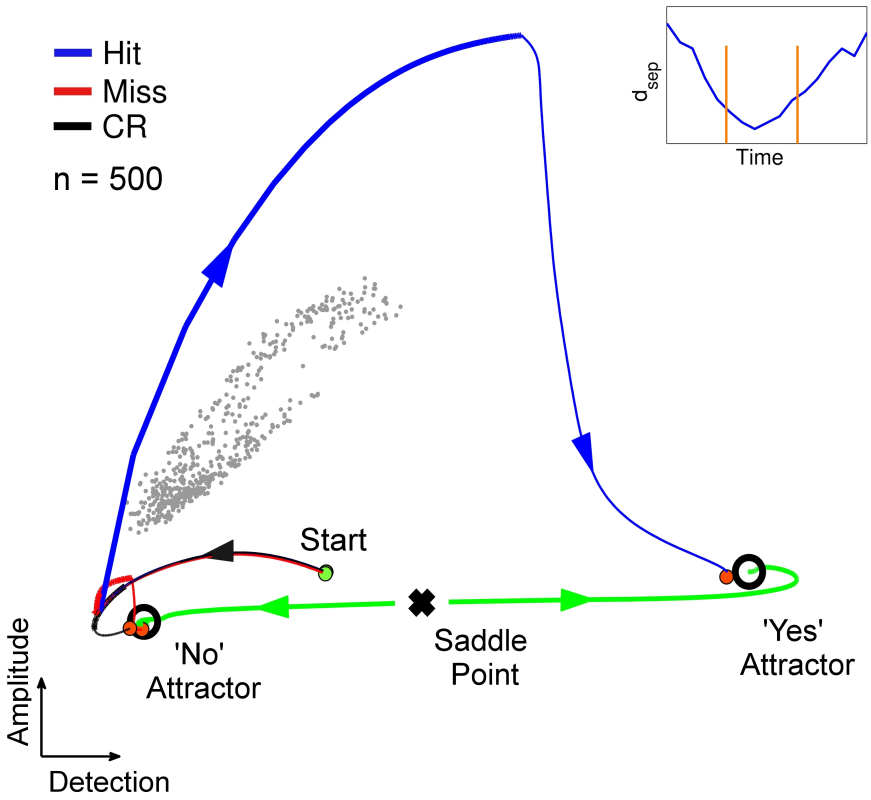

Figure 4.7: **Neural trajectories of the recurrent neural network model** Neural trajectories during a Hit (blue), a Miss (red) and a CR (black) trial projected in the same axes as in Figure 4. The three trajectories overlap during the beginning of the trial. The stimulus is applied (in the Hit and Miss conditions) at the middle of the possible stimulation window (thick black line in CR). The Hit trajectory evolves to the 'yes' attractor while the Miss and CR trajectories end in the 'no' attractor. The gray dots are points close to the separatrix, estimated as the states achieved during 'borderline' stimuli. Inset shows the distance between the network state during CR trials and the separatrix. The fixed points analysis of the trained network revealed a saddle point mediating the decision between the two stable fixed points. The green traces represent the trajectories starting near the saddle point following its unstable direction. For better visualization of this figure, the simulations were ran without noise in the sensory inputs, but the effects do not change under noisy stimuli.

4.3 Discussion

Under temporal uncertainty, the detection of a sensory stimulus embedded in a noisy background can be improved by previous knowledge about the probability of stimulus arrival over time. This improvement can arise from a dynamical modulation of the response criterion over time. We used novel analysis methods to extract the response criterion from neural data, demonstrating that it indeed is modulated according to the learned temporal structure of the task. Furthermore, we proposed a dynamical mechanism for the modulation of the response criterion and showed that PMc activity is consistent with it.

In the framework of Signal Detection Theory (Green et al. [1966]), the response criterion is expected to vary with the hazard rate - the probability of observing the stimulus in the next instant, knowing that it has not arrived up till now. For the uniform distribution of onset times used in this experiment, the hazard rate is increasing within the possible stimulation window. Although our estimation of the probability of false alarm over time from the experimental data is noisy (Figure 4.3), its profile does not seem to be consistent with a decreasing response criterion within the possible stimulation window, as predicted by the hazard rate. One possible explanation for this deviation from optimality is that monkeys might be able to identify periods in which the stimulus is more likely but not capable of estimating the exact distribution of stimulus onset times. Indeed, in a reaction-time task involving the detection of stimulus dysrhythmia with constant hazard rate, the FA rate over time was not constant but showed a peak at the beginning of the relevant temporal interval, (see Ghandehari [2000] figure 3.3).

The neural mechanism that we propose for modulating the response criterion over time relies on the network's recurrent dynamics. Because the subject is presented with many trials of different stimulus onset times, we speculate that information about the timing of the sensory evidence might be incorporated in the decision process by plastic changes in the internal synaptic connections of the 'decision' network (Janssen and Shadlen [2005]; Karmarkar and Buonomano [2007]; Leon and Shadlen [2003]). This builds on the framework of state-dependent or reservoir computing (Buonomano and Maass [2009]; Jaeger and Haas [2004];

Maass et al. [2007]) in which computations arise from the interaction between external stimuli and the internally generated dynamics produced by the network's recurrent connectivity. This framework is particularly suitable to explain neural mechanisms that require time- like the effect of temporal expectations- because arbitrary functions of time can be intrinsically encoded in reproducible neural trajectories (Laje and Buonomano [2013]). Notably, in our implementation, the modulation of the response criterion according to the time-varying probability of stimulation does not need to be explicitly trained. It arises from the multiple presentation of many trials of different stimulus onset times together with a target signal indicating the presence or absence of the stimulus. While the training algorithm is far from being biologically realistic (but see Hoerzer et al. [2014]), it is important to note that the information used during the online supervised learning was analogous to the one that the monkeys receive in the experimental setup.

The combination of previous knowledge about the stimulus probability with incoming sensory evidence was extensively studied in two-alternative forced choice discrimination tasks (Forstmann et al. [2010]; Hanks et al. [2011]; Rao et al. [2012]; Ratcliff and McKoon [2008]; Simen et al. [2009]; Summerfield and Koechlin [2008]). These studies suggest that stationary priors are incorporated into the decision process as a shift in the amount of evidence needed to reach a decision. In this work, we extended this question to the temporal domain. We used a detection task to ask how can subjects use prior information about the timing of stimulus arrival. Our results suggest a neural mechanism that supports the incorporation of a time-varying prior probability into the decision process.

How do our results extend to a discrimination task? We speculate that temporal expectations could dynamically shift the amount of required evidence for each choice according to their time-varying prior probability. In order to test this, one possibility would be to use a combined detection-discrimination task (Jaramillo and Zador [2011]) in which subjects must both detect and discriminate stimuli to receive reward. In this task, the relative frequency of each alternative could be manipulated so that it changes within the time course a trial. We anticipate that individuals will infer the task's temporal structure and use their temporal

expectations to dynamically modulate their bias for each alternative across time.

Note that the dynamic control of response criterion could be combined with the effect of other dynamic biases. In the reaction time version of the random dot task, for example, it has been shown that prior probabilities of the stimulus, even when stationary, could be incorporated as a dynamic bias signal, increasing the relative weight of priors over evidence as decision time increases (Hanks et al. [2011]). Thus, several effects could interact to shape the time-to-time amount of evidence required to reach a decision.

We devised a method to extract the timing of false alarm events from the neural activity based on template-matching of activity patterns. While a similar approach was previously used in the context of memory-trace replay during sleep (Louie and Wilson [2001]), here we apply it to extract decision-related information from neural activity. Our method is useful to provide timing information about the subject's decision in situations when this is not immediately reported by its behavior (i.e. when short-term memory of the chosen alternative is required). Although we developed it to infer a time-varying response criterion, our technique has a broader applicability. It could, for example, provide valuable insight into the sources of false alarms. If the activity of neurons in sensory cortices are recorded simultaneously with PMc neurons, our method could be used to provide the relevant times for building a 'false-alarm triggered averaged' of sensory activity. This could potentially disambiguate multiple possible origins of false alarms trials and contribute to the understanding of the role of noise on behavioral variability (Renart and Machens [2014]).

Subjects performing a decision-making task can benefit from the use of temporal expectations at multiple stages of the sensorimotor transformation. In this work, we showed that temporal information can be used to modulate the subject's response criterion across time. However, our experimental paradigm is not able to rule out other possibilities. For example, the sensory representation of stimuli could be changing over time (Correa et al. [2005]; Ghose and Bearl [2010]; Ghose and Maunsell [2002]; Jaramillo and Zador [2011]; Rohenkohl et al. [2012]). In the periods of higher expectations the signal to noise ratio of the sensory channel could be increased by mechanisms as synchronization (Steinmetz et al. [2000]).

During periods of lower expectation there could be gating mechanisms helping to avoid noise-induced false positives. Different experimental paradigms and further studies are needed to analyze the existence and coordination of different neural mechanisms for benefiting from temporal expectations.

4.4 Methods

4.4.1 Detection task

Data for this analysis was obtained from two earlier studies (de Lafuente and Romo [2005, 2006]). Stimuli were delivered to the skin of the distal segment of one digit of the restrained hand, via a computer-controlled stimulator (BME Systems; 2 mm round tip). Initial probe indentation was 500 μm. Vibrotactile stimuli consisted of trains of 20 Hz mechanical sinusoids with amplitudes of 2.3 - 34.6 mm. These were interleaved with an equal number of trials where no mechanical vibrations were delivered to the skin (amplitude = 0). Animals pressed one of two buttons to indicate stimulus present (left button) or stimulus absent (right button). They were rewarded with a drop of liquid both types of correct responses, i.e, correct detections in stimulus-present trials and correct rejections in stimulus-absent trials. Animals were handled in accordance with standards of the National Institutes of Health and Society for Neurosccience. All protocols were approved by the Institutional Animal Care and Use Committee of the Instituto de Fisiologa Celular

4.4.2 Recordings

Neuronal recordings were obtained with an array of seven independent, movable micro-electrodes (23 M Ω) inserted in the ventral premotor cortex (VPc), dorsal premotor cortex (DPc) and in medial premotor cortex (MPc) in both hemispheres. Neurons were selected if they responded to any of the different components of the detection task. The locations of the electrode penetrations were confirmed with

standard histological techniques. Cortical areas were identified based on cortical landmarks. Detailed description of the experimental techniques was described in de Lafuente and Romo [2005, 2006]. The experimental data set included 144 recording sessions from two monkeys (47 from monkey R16 and 97 from R19). Each session contained a variable number of simultaneously recorded neurons. The maximum number of simultaneous units was 6 and the median across sessions was 2. The total number of neurons was 384 (117 of monkey R16 and 267 in monkey R19).

4.4.3 Data analysis

4.4.3.1 FA detection by template-matching

For each neuron, we computed the firing rate using 250ms sliding windows displaced every 50ms. We considered the average over hit trials as the neuron-specific typical trajectory triggered by the vibratory stimulation. From this profile we selected a 1 s segment and used it as a template to find similar patterns in single FA trial. The template included the 0.5 s stimulation period and the first 0.5 s of delay period. We slided the template over single FA trials, computing, for each time, the mean squared error between the firing activity on the single trial and the template profile. Because of the 1 s width of the template, this error was defined from the beginning of the trial until 1 s before the end of the delay period. On each trial, a significant match was identified as a FA event if the error presented a minimum that exceeded 1.5 times the error's standard deviation over time. With this algorithm, we found that 347 out of the 384 recorded neurons had at least one FA trial with a FA event and in approximately 80% of the neurons more that 75% of the FA trials contained a FA event.

To test the significance of the detected events we used the activity of simultaneously recoded neurons. We independently detected events on each trial from the activity of the two different neurons. If an event corresponds to a false percept, it should be detected at the same time on simultaneously recorded neurons. We computed the frequency of differences in the detected times, and compared

it to both chance level and CR trials. Chance level was obtained by shuffling the trials, keeping the same set of detected times but breaking the trial-to-trial correspondence between neurons. The significance of simultaneous detections (within 350 ms, first bin in the histogram of Figure 4.2d) was tested with a *z-test* resulting in a $p<0.001$.

The probability of producing a FA over time was estimated as the number of trials in which of FA event was detected in 500 ms temporal windows, normalized by the total number of FA trials. We corrected for the different trials durations by considering those trials that ended within a time bin as contributing as a fraction to the normalization term.

4.4.3.2 State-space analysis

We constructed pseudo-simultaneous population responses by combining neural data mostly recorded separately. Matching the conditions between different recording sessions resulted in $N = 132$ neurons from which we had data in every condition (Hits and Misses of several amplitudes, CR's and FA's). We projected the averaged activity of these neurons onto two task-related axes: stimulus amplitude ($\mathbf{a}_{amp}$) and stimulus detection ($\mathbf{a}_{det}$).

The stimulus amplitude axis, $\mathbf{a}_{amp}$, was obtained as the set of coefficients that best relate each neuron's trial-to-trial response to the stimulus amplitude. To find it, we used a multivariate regression analysis on the firing rate r of each neuron k following

$$r_i^k(t) = \beta_1^k(t)\ amp_i + \beta_2^k(t)\ choice_i + \beta_3^k(t)\ amp_i\ choice_i + \beta_4^k(t) \tag{4.1}$$

where $amp(i)$ and $choice(i)$ denote the stimulus amplitude and the subject's choice in trial i, respectively. The stimulus amplitude axis $\mathbf{a}_{amp}$ was defined as the set of coefficients β_1^k for the N recorded neurons ($k = 1...N$) at the stimulus onset time. The firing rate r was calculated using bins of 100 ms, so they were

large enough to include the effect of the stimulus.

$$\mathbf{a}_{amp} = \left[\beta_1^1(t_{SO})\ \beta_1^2(t_{SO})\ ...\ \beta_1^N(t_{SO})\right]' \tag{4.2}$$

This axis represents the direction in neural space in which the stimulus drives the network.

The stimulus detection axis, $\mathbf{a}_{det}$, was defined as the the vector connecting the population activity just before the application of the stimulus to the that at the end of the delay period, during hit trials.

$$\mathbf{a}_{det} = \mathbf{r}_H(t_{ED}) - \mathbf{r}_H(t_{SO} - \Delta t) \tag{4.3}$$

where $\mathbf{r}_H$ is the N-dimensional vector of neural activity averaged over hit trials. The stimulus detection axis, $\mathbf{a}_{det}$, represents the direction in which the network evolves when the subject detects a stimulus.

4.4.4 Recurrent network model

We used a recurrent network of $N = 500$ nonlinear firing-rate units. Each unit is described by an activation variable x_i evolving as,

$$\tau\frac{dx_i}{dt} = -x_i + g\sum_{j=1}^{N} J_{ij}r_j + w_i^{fb}z + w_i^{start}u_{start} + w_i^{stim}u_{stim} \tag{4.4}$$

where $r_i = \tanh(x_i)$ is the 'firing rate' and $z = \sum_{i=1}^{N} w_i^{out}r_i$ is the network's output. The sparse matrix $\mathbf{J}$ stores the recurrent connection weights and had density $p = 0.1$, meaning that each element had probability $1 - p$ of being set to 0. The nonzero elements of $\mathbf{J}$ were drawn from a Gaussian distribution with mean zero and variance $1/Np$. The parameter g that scales the strengths of the recurrent connections was set to 1.2. The neuronal time constant is $\tau = 100$ ms and the simulations were performed by Euler integration with a step of $dt =$

$10ms$. The network received two external inputs, u_{start} and u_{stim}, representing the start cue and the sensory channel, respectively. Each neuron received the inputs through a randomly chosen synaptic strength, w_i^{start} and w_i^{stim}. The start cue, u_{start}, is a 500 ms pulse applied at the beginning of each trial. The sensory input, u_{stim}, is modeled as a 300 ms pulsed signal proportional to the stimulus amplitude and embedded in a noisy background. The sensory noise was produced by an Ornstein-Uhlenbeck process of variance $\sigma^2 = 0.4$ and correlation time $\tau = 0.3s$. The decision was extracted from a linear readout of the network activity, z, whose coefficients w^{out} are initially set to zero and then modified by the learning algorithm.

Trials begin with the start cue signal, and after a variable pre-stimulation period, on half of the trials, the stimulus is applied. Trials of different stimulus amplitudes were randomly interleaved with stimulus-absent trials. The pre-stimulation period varies from trial to trial, taking values homogeneously distributed between 0.6 s and 1 s. The stimulus has a duration of 300 ms in every trial, even when its amplitude is 0 (stimulus-absent trials). After the offset of the stimulus there is a delay period which has a fixed duration of 400 ms. The valid decision interval is a 500 s temporal window that starts after the end of delay period.

Training was performed using the FORCE algorithm (Sussillo and Abbott [2009]) to modify the output weights w^{out}. Although these are the only plastic weights, the feedback weights $\mathbf{w}^{fb}$ translate this into a rank-one perturbation to the effective recurrent weight matrix $\mathbf{J}_{\text{eff}} = \left(g\mathbf{J} + \mathbf{w}^{fb}\mathbf{w}^{out'}\right)$ (Sussillo and Abbott [2012]). The desired output during training trials was zero at all times, except during the decision interval in which it was 1 if the stimulus was present or 0 if it was absent.

Once trained we quantified the performance of the network through the 'psychometric' function. We simulated the trained network for 2000 trials and obtained the frequency of stimulus-present responses as a function of the stimulus amplitude. To estimate the network's response criterion, we systematically probed the network with variable amplitude stimuli at different times from the beginning of the trial. We followed a bisection protocol for the stimulus ampli-

tudes to look for the lowest amplitude that led to a stimulus-present response. Measuring this borderline stimulus at different times from the start cue gives an estimate of the response criterion over time. We also obtained the rate of FA as a function of time by using the template-matching algorithm described above. We applied to the model the same algorithm that we used for the experimental data.

4.4.4.1 Fixed points analysis

To analyze the dynamics of the trained network we used the technique developed in Sussillo and Barak [2013]. The network defined in Eq 4.4 is a high-dimensional dynamical system. To understand its behavior, we looked for fixed and slow points of phase space by minimizing the function

$$q(\mathbf{x}) = \frac{1}{2}|\mathbf{f}(\mathbf{x})|^2 \tag{4.5}$$

where

$$\mathbf{f}(\mathbf{x}) = -\mathbf{x} + \mathbf{J}^{\text{eff}} \tanh(\mathbf{x}) \tag{4.6}$$

and $\mathbf{J}_{\text{eff}} = \left(g\mathbf{J} + \mathbf{w}^{fb}\mathbf{w}^{out'}\right)$. The vector function $\mathbf{f}(\mathbf{x})$ defines the nonlinear dynamical system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, presented in Eq 4.4.

In order to find minimums of $q(\mathbf{x})$, we simulated the model with several stimulus amplitudes and used the state of the network ($\mathbf{x}$) at different points in time as initial conditions for the minimization algorithm. This procedure systematically identified 3 relevant fixed points. Then, for each point $\mathbf{x}^*$ we defined the local linear approximation, $\dot{\delta\mathbf{x}} = \mathbf{M}\delta\mathbf{x}$, where

$$M_{ij} = \frac{\delta f_i}{\delta x_j} = -\delta_{ij} + J_{ij}^{\text{eff}} \left[1 - \tanh(x_j)^2\right]. \tag{4.7}$$

By studying the eigenvalues of $\mathbf{M}$ we analyzed the stability of each fixed point.

Chapter 5

Final conclusions

Central to cognitive neuroscience is to understand how perception and decision-making arise from neural activity. In this thesis, I aimed to explore this link in the context of the . Three essential issues have been addressed: (1) the dynamics of correlated variability, (2) the decoding of a decision from neural population's activity and (3) the neural mechanisms underlying the use of temporal expectations. Importantly, all these issues have been studied within the same experimental paradigm.

The importance of correlated variability is, at least, twofold. On the one hand, correlations between the activity of neurons can be used as a tool to understand neural circuits (just as any statistical measure of firing activity, De La Rocha et al. [2007]; Renart et al. [2010]). In the vibrotactile detection task, the dynamics of choice-conditioned noise correlations reveal an internal component of the decision-making process (Chapter 2) that relates to the temporal uncertainty and motivates the study in Chapter 4. On the other hand, correlated variability is fundamental to understand information encoding: neural activity is significantly noisy and the impact of this noise has been shown to critically depends on its correlation among neurons (Averbeck et al. [2006]; Cohen and Newsome [2009]; Haefner et al. [2013]; Nienborg and Cumming [2010]; Nienborg et al. [2012]). Choice-conditioned noise correlations in the vibrotactile detection task are weak, specially during the period in which subjects are required to hold their choice

in short-term memory. The impact of noise correlation on the encoding (and decoding) of a perceptual decision that has to be stored in working memory has motivated the work in Chapter 3.

Decisions emerge from the concerted activity of neuronal populations. However, the analytical tools best suited to decode decision signals from neuronal populations are unclear. In this thesis, I developed and applied statistical measures to quantify how predictive of the subject's decision is the activity of populations of premotor cortex neurons. I found that, while populations of S2 cortex and sensory-like PMc neurons are only partially correlated with behavior, those PMc neurons active during a delay period preceding the motor report predict unequivocally the animal's decision report.

Finally, the vibrotactile detection task requires subject's to make a decision about the presence of a stimulus whose arrival time is unknown. How does a neural system cope with this temporal uncertainty? In Chapter 2, I showed that the dynamics of firing rates and noise correlations suggest the existence of an internal component of the decision-making process related with the task's temporal structure. In Chapter 4, I followed this evidence and found that subjects build and benefit from temporal expectations by modulating their response criterion over time. Moreover, I showed that this modulation is represented by the population activity of PMc neurons. Using a recurrent neural model, I proposed a novel mechanism to implement this modulation. Knowledge about the stimulus probability over time, acquired during training, is intrinsically encoded in the neural population dynamics.

The study of the neural basis of perception and perceptual decision-making is relatively new but significant progress has been achieved rapidly (Gold and Shadlen [2007]; Romo and de Lafuente [2013]). However, key issues still remain elusive: what is the role of behavior variability? How does it relate with neural variability? How does neural noise impact our brain's information encoding and transmission capability? Where are the limits of perception? While a large amount of theoretical effort has been spent on trying to elucidate these problems, it is only now that technological advances let us test them experimentally and in a straightforward manner. The combination of psychophysical and neurophysi-

ological experiments with optogenetic perturbative and observational tools sets an unprecedented scenario to answer these long-standing questions (Fenno et al. [2011]; Stosiek et al. [2003]). These are early days, but the rapid development of these new technologies capable of causally probe entire neural circuits suggest that a mechanistic answer of how our perceptions and decision arise from neural activity is not that far.

Chapter 6

Conclusiones Finales

Entender cómo la percepción y la toma de decisiones emergen de la actividad neuronal es un objetivo central de la neurociencia cognitiva. En esta tesis exploré este asunto en el contexto de la detección de estímulos vibrotáctiles. He abordado tres asuntos esenciales: (1) la dinámica de la variabilidad correlacionada, (2) la decodificación de una decisión a partir de la actividad de una población de neuronas y (3) los mecanismos neurales que subyacen a las expectativas temporales.

La importancia de la variabilidad correlacionada es, al menos, doble. En primer lugar, la correlación entre la actividad de pares de neuronas puede ser utilizada como herramienta para entender el funcionamiento de los circuitos neurales (así como cualquier otra medida estadística de la actividad neuronal, De La Rocha et al. [2007]; Renart et al. [2010]). En la tarea de detección de estímulos vibrotáctiles, la dinámica de las correlaciones de ruido, condicionadas a la decisión del animal, revelan una componente interna del proceso de decisión (Capítulo 2) relacionado con la estructura temporal de la tarea y que motiva el estudio del Capítulo 4. Por otro lado, la variabilidad correlacionada es fundamental para entender la codificación de información: la actividad neuronal es significativamente variable y el impacto de este ruido depende críticamente en la correlación entre neuronas (Averbeck et al. [2006]; Cohen and Newsome [2009]; Haefner et al. [2013]; Nienborg and Cumming [2010]; Nienborg et al. [2012]). Las correlaciones de ruido en la tarea de detección son débiles, en especial durante el período du-

rante el cual los sujetos deben mantener su decisión en memoria de trabajo. El impacto de las correlaciones del ruido en la codificación y decodificación de la decisión durante ese período motivó el trabajo descripto en el Capítulo 3.

Las decisiones emergen de patrones de actividad de poblaciones de neuronas. Sin embargo, las herramientas analíticas más adecuadas para estudiar decisiones perceptuales a partir de poblaciones de neuronas no son del todo conocidas. En esta tesis, he desarrollado y aplicado medidas estadísticas para cuantificar cuán predictiva sobre la decisión del animal es la actividad de neuronas en PMc. Como resultado, encontré que, mientras que poblaciones de neuronas en S2 y neuronas del tipo sensorial en PMc están parcialmente correlacionadas con el comportamiento, aquellas neuronas de PMc que están activas durante el período que antecede al movimiento, predicen inequívocamente la decisión del animal.

Finalmente, la tarea de detección de estímulos vibrotáctiles require que los sujetos tomen una decisión acerca de la presencia de un estímulo cuyo tiempo de llegada es ambiguo. ¿Cómo puede un sistema neuronal lidiar con esta incertidumbre temporal? En el Capítulo 2, mostré como la dinámica de la tasa de disparo y de la variabilidad correlacionada sugieren la presencia de una componente interna del proceso de toma de decisión. En el Capítulo 4, profundicé en este asunto y encontré que los sujetos construyen y utilizan expectativas temporales, modulando su criterio de respuesta a lo largo del tiempo del ensayo. Esta modulación se encuentra representada en la actividad poblacional de las neuronas de PMc. Utilizando un modelo de red recurrente, propuse un nuevo mecanismo mediante el cual se implementa esta modulación. El conocimiento acerca de la probabilidad de estimulación en función del tiempo, adquirido durante el entrenamiento, puede ser intrínsicamente codificado en la dinámica de la población neuronal.

El estudio de las bases neuronales de la percepción y la toma de decisiones perceptuales, aunque relativamente nuevo, ha progresado rápidamente (Gold and Shadlen [2007]; Romo and de Lafuente [2013]). Sin embargo, hay elementos claves que siguen siendo desconocidos: cómo se explica la variabilidad en la decisiones perceptuales de un sujeto durante múltiples repeticiones de la misma tarea? Cómo se relaciona esta variabilidad en el comportamiento con la variabilidad en la actividad eléctrica de las neuronas? Cuál es el impacto del ruido neuronal

en la capacidad de nuestro cerebro de codificar información? Dónde están los límites de la percepción? Grandes esfuerzos teóricos se han invertido en intentar resolver estos interrogantes. Sin embargo, recién ahora los avances tecnológicos permiten abordar estos temas experimentalmente. La combinación de experimentos psicofísicos y neurofisiológicos con herramientas optogenéticas representa un escenario sin precedentes para atacar estas preguntas (Fenno et al. [2011]; Stosiek et al. [2003]). El rápido avance de estas nuevas tecnologías, capaces de examinar circuitos neuronales con altísima resolución espacial y temporal, sugiere que una respuesta mecanística de como la percepción y las decisiones perceptuales emergen de la actividad neuronal no esta muy lejos.

Appendix A

Supplemental Material to Chapter 3

A.1 $CP_{2,w}$ index: verification of the analytical approximation

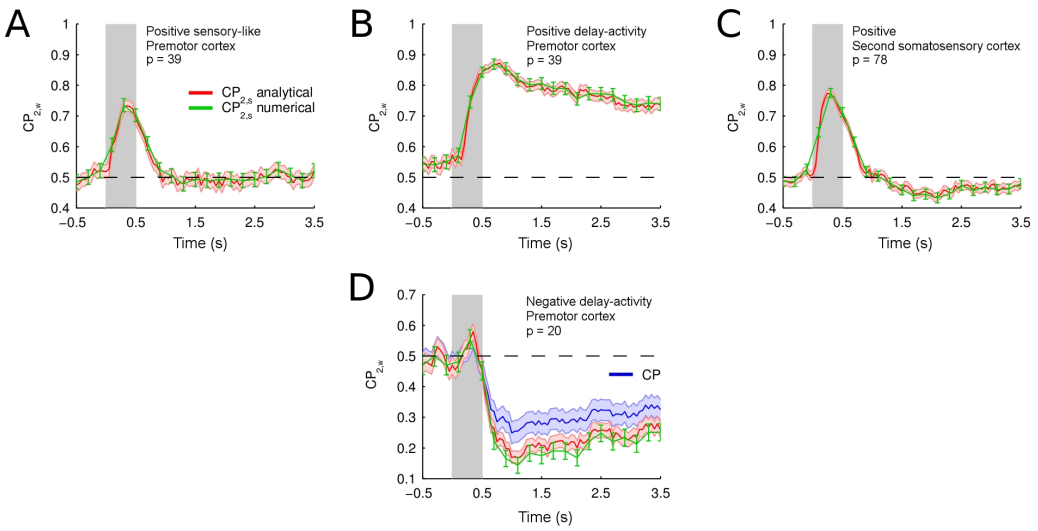

Figure A.1: **Numerical and analytical computation of the $CP_{2,w}$ index.** Related to Figure 2. Population-averaged $CP_{2,w}$ index computed with Equation A.16 (red traces) compared with the one obtained by direct numerical evaluation (green traces, see Methods). Gray boxes indicate the period of stimulation and p is the number of neurons used in this analysis. **(A)** Pool of positive sensory-like PM neurons. **(B)** Pool of positive delay-activity PM neurons. **(C)** Pool of positive S2 neurons. **(D)** Pool of negative delay-activity PM neurons. For this case we also show the population-averaged CP index in order to verify that the sum of activity of two neurons is more predictive than the activity of single neurons. Due to limitations in the number of simultaneously recorded neural pairs in our database, we cannot perform this analysis in pools of negative sensory-like PM and negative S2 neurons.

A.2 Further data on noise correlations and CP

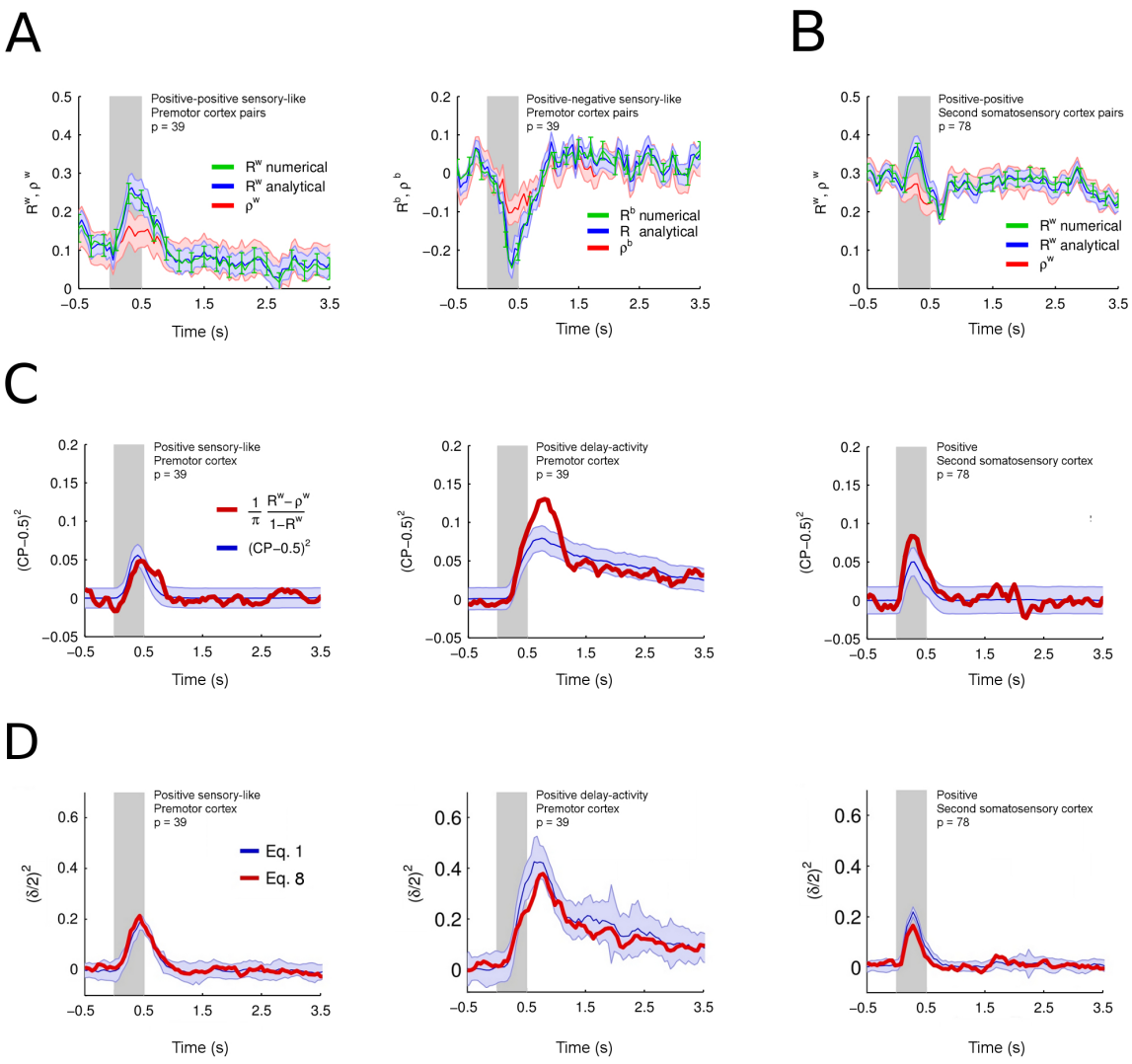

Figure A.2: **Noise correlations and choice-conditioned noise correlations determine the CP index.** Related to Figure 3. **(A)-(B)** Temporal evolution of the mean correlation coefficients computed with all trials (R,blue and green traces) compared with the average of correlations obtained using hit and miss trials separately (ρ, red traces). Mean correlation coefficients were obtained averaging over all pairs from the same functional type. Gray boxes indicate the period of stimulus presentation, error bars represent standard error of the mean (SEM) and p is the number of pairs. Green traces are the correlation coefficients computed numerically. Blue traces are the predictions from Equation A.48. **(A)** Pairs within the pool of positive sensory-like PM neurons (Left) and pairs of positive and negative PM sensory-like neurons (Right). **(B)** Pairs within the pool of positive S2 neurons. **(C)** Verification of the linear approximation for the CP in terms of the correlation coefficients (Equation A.55). (Left) Pool of positive sensory-like PM neurons. (Middle) Pool of positive delay-activity PM neurons. (Right) Pool of positive S2 neurons. As expected the agreement is good except for large values of the CP index. **(D)** Population averaged $\left(\frac{\delta_0}{2}\right)^2$ derived from Equation 3.1 (blue traces) compared with the one obtained using Equation 3.8 (red traces). Gray boxes indicate the stimulation period and p the number of pairs. (Left) Pool of positive sensory-like PM neurons. (Middle) Pool of positive delay-activity PM neurons. (Right) Pool of positive S2 neurons.

A.3 $CP_{2,b}$ index and optimal rate combination for pairs of S2 neurons

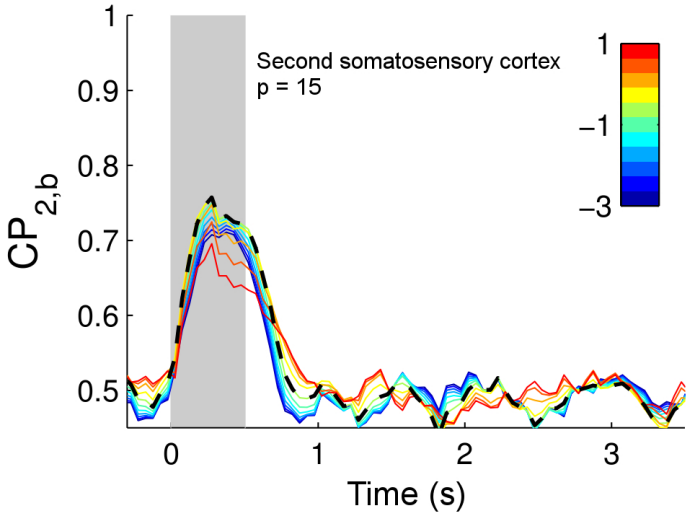

Figure A.3: **Optimal rate combination for pairs of S2 neurons**Related to Figure 4. Population-averaged $CP_{2,b}$ for pairs of positive and negative S2 neurons, computed using Equation A.22 for different values of D (Color coded). The rate combination with optimal $CP_{2,b}$, during the stimulation period, is obtained for $D = -0.7$ (black dotted line).

A.4 Dependence of the CPN index on the number of neurons

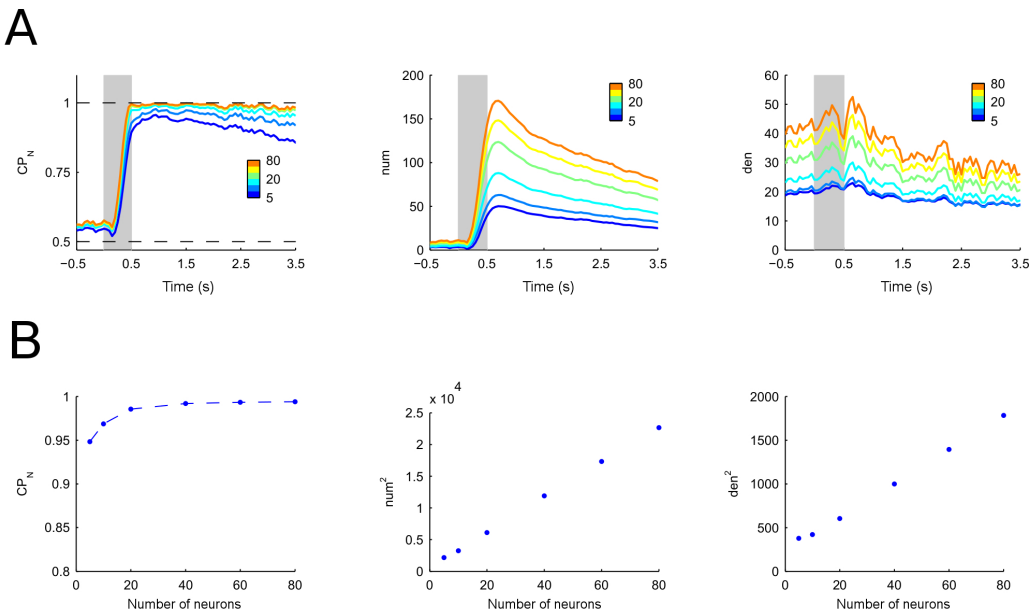

Figure A.4: **Dependence of the CP_N index on the number of neurons.** Related to Figure 5. **(A)** Time course of the CP_N index (Left), numerator (Middle) and denominator (Right) of Equation A.61 for N taking values between 5 and 80 neurons per pool (color-coded). For small N the CP_N index increases with N while at larger values of N it saturates. **(B)** Dependence on N of the CP_N index (Left), the square of the numerator of Equation A.61 (Middle) and the square of the denominator (Right) for a fixed time, t = 1 s.

A.5 Choice Probability Indices

The generalized choice probability index of an arbitrary variable R is defined as the area under the receiver-operating-characteristic (ROC) curve, analogously to the choice probability index. It can be expressed as,

$$\text{GCP} = \int_0^1 \beta d\alpha \tag{A.1}$$

where,

$$\begin{aligned} \alpha(z) &= P\left[R > z|B\right] \\ \beta(z) &= P\left[R > z|A\right] \end{aligned} \tag{A.2}$$

A and B are the two possible choices and z is the threshold level. The quantity $\alpha(z)$ is the probability of finding a type B trial with a value of R higher than the threshold z. In the same way, $\beta(z)$ is the probability of finding a type A trial with R higher than the threshold. In Equation A.1 we are assuming that the mean value of thedistribution of R over A trials is higher than the one in B trials, but the opposite case is analogous. The probability that a B trial is classified as A, as a function of the threshold z, can be expressed as,

$$\alpha(z) = \int_z^\infty P\left[R|B\right] dR \tag{A.3}$$

from where,

$$\frac{d\alpha}{dz} = -p[z|B]. \tag{A.4}$$

Then, Equation A.1 can be written as,

$$\text{GCP} = \int_{-\infty}^{\infty} p[z|B]\ \beta(z)\ dz. \tag{A.5}$$

Note that, because R is an arbritrary variable, it can take any value in the real axis. Therefore, the threshold z can also have any value from $-\infty$ to ∞. If we assume that R has a gaussian distribution in each of the two set of trials, then we can write,

$$p[z|B] = \frac{1}{\sqrt{2\pi\Sigma_B^2}} \exp\left(-\frac{(z-\Theta_B)^2}{2\Sigma_B^2}\right) \tag{A.6}$$

and

$$\begin{aligned} \beta(z) &= p[R > z|A] \\ &= \frac{1}{2}\left[1 - \text{erf}\left(\frac{z-\Theta_A}{\sqrt{2\Sigma_A^2}}\right)\right] \\ &= \frac{1}{2}\,\text{erfc}\left[\frac{z-\Theta_A}{\sqrt{2\Sigma_A^2}}\right] \end{aligned} \tag{A.7}$$

where,

$$\begin{aligned} \Theta_A &= \langle R \rangle_A \\ \Theta_B &= \langle R \rangle_B \\ \Sigma_A^2 &= \langle (R-\Theta_A)^2 \rangle_A \\ \Sigma_B^2 &= \langle (R-\Theta_B)^2 \rangle_B \end{aligned} \tag{A.8}$$

Replacing both expressions in Equation A.5, we can express GCP as,

$$\text{GCP} = \frac{1}{2}\text{erfc}\left(-\frac{\Delta}{2}\right), \quad \Delta = \frac{\Theta_A - \Theta_B}{\sqrt{\frac{1}{2}\left(\Sigma_A^2 + \Sigma_B^2\right)}} \tag{A.9}$$

A.5.1 Single neuron firing activity

When the arbitrary variable R is the firing rate r of a neuron, the generalized choice probability reduces to the choice probability index and can be expressed analytically as,

$$\mathrm{CP} = \frac{1}{2}\mathrm{erfc}\left(-\frac{\delta}{2}\right), \quad \delta = \frac{\mu_A - \mu_B}{\sqrt{\frac{1}{2}\left(\sigma_A^2 + \sigma_B^2\right)}} \tag{A.10}$$

where,

$$\begin{aligned} \mu^A &= \langle r \rangle_A \\ \mu^B &= \langle r \rangle_B \\ \sigma_A^2 &= \langle (r - \mu_A)^2 \rangle_A \\ \sigma_B^2 &= \langle (r - \mu_B)^2 \rangle_B \end{aligned} \tag{A.11}$$

that corresponds to Equation 1 of the main text.

A.5.2 Sum of firing activities of two neurons ($\mathrm{CP}_{2,w}$)

As the simplest example of GCP index, consider the case where R is the sum of firing rates of two neurons from the same neural pool, $r_w = r_1 + r_2$.

$$\begin{aligned} \Delta_{2,w} &= \frac{\langle r_w \rangle_A - \langle r_w \rangle_B}{\sqrt{\frac{1}{2}\left(\Sigma_A^2 + \Sigma_B^2\right)}} \\ &= \frac{\left(\langle r_1 \rangle_A + \langle r_2 \rangle_A\right) - \left(\langle r_1 \rangle_B + \langle r_2 \rangle_B\right)}{\sqrt{\frac{1}{2}\left(\Sigma_A^2 + \Sigma_B^2\right)}} \end{aligned} \tag{A.12}$$

The variance Σ_c^2, for $c = A, B$ can be expressed as,

$$\begin{aligned}\Sigma_c^2 &= \langle r_w^2 \rangle_c - \langle r_w \rangle_c^2 \\ &= \sigma_{1,c}^2 + \sigma_{2,c}^2 + 2\sigma_{1,c}\sigma_{2,c}\rho^{w,c}\end{aligned} \tag{A.13}$$

where,

$$\begin{aligned}\sigma_{1c}^2 &= \langle r_1^2 \rangle_c - \langle r_1 \rangle_c^2 \\ \sigma_{2c}^2 &= \langle r_2^2 \rangle_c - \langle r_2 \rangle_c^2\end{aligned} \tag{A.14}$$

and $\gamma_{12}^c = \langle r_1 r_2 \rangle_c - \langle r_1 \rangle_c \langle r_2 \rangle_c$ is the covariance between r_1 and r_2 evaluated over trials of type $c = A, B$.

From this we have,

$$\mathrm{CP}_{2,w} = \frac{1}{2}\mathrm{erfc}\left(-\frac{1}{2}\Delta_{2,w}\right). \tag{A.15}$$

where

$$\Delta_{2,w} = \frac{\left(\mu_1^A - \mu_1^B\right) + \left(\mu_2^A - \mu_2^B\right)}{\sqrt{\frac{1}{2}\left(\sigma_{1,A}^2 + \sigma_{1,B}^2\right) + \frac{1}{2}\left(\sigma_{2,A}^2 + \sigma_{2,B}^2\right) + \left(\gamma_{12}^A + \gamma_{12}^B\right)}} \tag{A.16}$$

A simple particular case is obtained if the variance of the firing rate distribution is equal for both neurons and both types of trials. Denoting this variance as σ^2 and defining the correlation $rho_{12}^{w,c} = \gamma_{12}^c/\sigma^2$, we have

$$\mathrm{CP}_{2,w} = \frac{1}{2}\mathrm{erfc}\left(-\frac{1}{2}\Delta_{2,w}\right). \tag{A.17}$$

$$\Delta_{2,w} = \frac{\left(\mu_1^A - \mu_1^B\right) + \left(\mu_2^A - \mu_2^B\right)}{\sigma\sqrt{2 + \left(\rho^{w,A} + \rho^{w,B}\right)}} = \frac{\sqrt{2}\delta_{1,2}}{\sqrt{1 + \rho_{12}^w}} \tag{A.18}$$

where $\delta_{1,2}$ is the arithmetic mean of δ_1 and δ_2, and $\rho_{12}^w = 0.5\left(\rho_{12}^{w,A} + \rho_{12}^{w,B}\right)$. The superscript w indicates that neurons 1 and 2 are within the same neural pool.

A.5.3 Arbitrary combination of the firing activity of two neurons ($\mathbf{CP}_{2,b}$)

If the two neurons belong to different pools, we take R as an arbitrary linear combination of their firing rates, $r_b = C_1\ r_1 + C_2\ r_2$.

$$\begin{aligned}\Delta_{2,b} &= \frac{\langle r_b\rangle_A - \langle r_b\rangle_B}{\sqrt{\frac{1}{2}\left(\Sigma_A^2 + \Sigma_B^2\right)}} \\ &= \frac{C_1\left(\langle r_1\rangle_A - \langle r_1\rangle_B\right) + C_2\left(\langle r_2\rangle_A - \langle r_2\rangle_B\right)}{\sqrt{\frac{1}{2}\left(\Sigma_A^2 + \Sigma_B^2\right)}}\end{aligned} \tag{A.19}$$

The variance of r_b in trials of type $c = A, B$ is,

$$\Sigma_c^2 = C_1^2\ \sigma_{1,c}^2 + C_2^2\ \sigma_{2,c}^2 + 2\ C_1\ C_2\ \gamma_{12}^c \tag{A.20}$$

From this we have,

$$\mathrm{CP}_{2,b} = \frac{1}{2}\mathrm{erfc}\left(-\frac{1}{2}\Delta_{2,b}\right). \tag{A.21}$$

$$\Delta_{2,b} = \frac{\left(\mu_1^A - \mu_1^B\right) + d\left(\mu_2^A - \mu_2^B\right)}{\sqrt{\frac{1}{2}\left(\sigma_{1,A}^2 + \sigma_{1,B}^2\right) + \frac{1}{2}D^2\left(\sigma_{2,A}^2 + \sigma_{2,B}^2\right) + D\left(\gamma_{12}^A + \gamma_{12}^B\right)}} \tag{A.22}$$

where we have defined $D = C_2/C_1$.

Again, a simple particular case is obtained if the variance of the firing rate distribution is equal for both neurons and both types of trials. Denoting this variance as σ^2 and defining $\rho_{12}^{b,c} = \gamma_{12}^c/\sigma^2$ we have

$$CP_{2,b} = \frac{1}{2}\text{erfc}\left(-\frac{1}{2}\Delta_{2,b}\right) \tag{A.23}$$

$$\Delta_{2,b} = \frac{\left(\mu_1^A - \mu_1^B\right) + D\left(\mu_2^A - \mu_2^B\right)}{\sigma\sqrt{1 + D^2 + D\left(\rho^{b,A} + \rho^{b,B}\right)}} = \frac{\delta_1 + D\delta_2}{\sqrt{1 + D^2 + D\rho_{12}^b}} \tag{A.24}$$

where $\rho_{12}^b = 0.5\left(\rho_{12}^{b,A} + \rho_{12}^{b,B}\right)$. The superscript b indicates that neurons 1 and 2 belong to different neural pools.

A.5.4 Arbitrary combination of the firing activity of two neural pools ($\mathbf{CP_N}$)

Consider now the case of two neural pools, denoted by subscripts + and -, having N_+ and N_- neurons respectively. Let us take R as a linear combination of their mean firing activities,

$$r_{\text{N}} = C_1 \sum_{j=1}^{N_+} r_{+,j} + C_2 \sum_{j=1}^{N_-} r_{-,j} \tag{A.25}$$

where $r_{\alpha,j}$ is the firing rate of neuron j in population $\alpha = +, -$. In order to calculate the ROC index, we have to obtain,

$$\Delta_{\mathrm{N}} = \frac{\langle r_N \rangle_A - \langle r_N \rangle_B}{\sqrt{\frac{1}{2}\left(\Sigma_A^2 + \Sigma_B^2\right)}} \tag{A.26}$$

In this case,

$$\Delta_{\mathrm{N}} = \frac{C_1 N_+ \left(\bar{\mu}_+^A - \bar{\mu}_+^B\right) + C_2 N_+ \left(\bar{\mu}_-^A - \bar{\mu}_-^B\right)}{\sqrt{\frac{1}{2}\left(\Sigma_A^2 + \Sigma_B^2\right)}} \tag{A.27}$$

where the bar indicates an average over the population, i. e.,

$$\bar{\mu}_+^c = \frac{1}{N_+} \sum_{j=1}^{N_+} \mu_{+j} \tag{A.28}$$

$$\bar{\mu}_-^c = \frac{1}{N_-} \sum_{j=1}^{N_-} \mu_{-j} \tag{A.29}$$

The variance Σ_c^2 is now expressed as,

$$\begin{aligned} \Sigma_c^2 &= \langle r_N^2 \rangle_c - \langle r_N \rangle_c^2 \\ &= C_1^2 \left(N_+ \bar{\sigma}_{+,c}^2 + N_+(N_+ - 1)\bar{\gamma}_{++}^c\right) + \\ &+ C_2^2 \left(N_- \bar{\sigma}_{-,c}^2 + N_-(N_- - 1)\bar{\gamma}_{--}^c\right) + \\ &+ 2C_1 C_2 N_+ N_- \bar{\gamma}_{+-}^c \end{aligned} \tag{A.30}$$

Using $D = C_2/C_1$, we have,

$$\mathrm{CP_N} = \frac{1}{2}\mathrm{erfc}\left(-\frac{\Delta_N}{2}\right) \tag{A.31}$$

$$\Delta_{\mathrm{N}} = \frac{N_+ \left(\bar{\mu}_+^A - \bar{\mu}_+^B\right) + DN_- \left(\bar{\mu}_-^A - \bar{\mu}_-^B\right)}{\sqrt{\frac{1}{2}\left(\frac{\Sigma_A^2}{C_1^2} + \frac{\Sigma_B^2}{C_1^2}\right)}} \tag{A.32}$$

where,

$$\begin{aligned}\frac{\Sigma_c^2}{C_1^2} &= N_+\bar{\sigma}_{+,c}^2 + N_+(N_+ - 1)\bar{\gamma}_{++}^c + \\ &+ D^2\left(N_-\bar{\sigma}_{-,c}^2 + N_-(N_+ - 1)\bar{\gamma}_{--}^c\right) + \\ &+ 2DN_+N_-\bar{\gamma}_{+-}^c\end{aligned} \tag{A.33}$$

In the application of Equations A.31-A.33 to experimental data, we estimated the population-averaged quantities pooling neurons and pairs across different recording sessions. The population-averaged mean and variance of the firing rate in condition c, denoted by $\bar{\mu}_i^c$ and $\bar{\sigma}_{i,c}^2$ respectively, were obtained by pooling neurons from population $i = +, -$ recorded across all the experimental sessions. To estimate the population averaged covariances of firing rates in condition c, referred as $\bar{\gamma}_{ij}^c$, we averaged over all the recorded pairs consisting of one neuron belonging to population $i = +, -$ and the other to population $j = +, -$ across all recording sessions. In general, the $\mathrm{CP_N}$ index will depend on the number of neurons. In Figure A.4 we explore this dependence in our experimental data.

A.5.5 Weighted sum of the activity of a pool of neurons

In a more general context one can study the covariation with behavior of arbitrary linear combinations of the neurons firing rates. We want to study how the variable

$$r_N = \sum_{i=1}^{N} w_i r_i = \boldsymbol{w}^T \boldsymbol{r} \tag{A.34}$$

convaries with behavior. The generalized choice probability reads,

$$\mathrm{CP_N} = \frac{1}{2}\mathrm{erfc}\left(-\frac{\Delta_N}{2}\right), \quad \Delta_\mathrm{N} = \frac{\mu_N^A - \mu_N^B}{\sqrt{\frac{1}{2}\left(\sigma_{N,A}^2 + \sigma_{N,B}^2\right)}} = \frac{\boldsymbol{w}^T\left(\boldsymbol{\mu}^A - \boldsymbol{\mu}^B\right)}{\sqrt{\boldsymbol{w}^T\boldsymbol{\Gamma}\boldsymbol{w}}} \quad \text{(A.35)}$$

where

$$\begin{aligned}\mu_N^c &= \left\langle \sum_{i=1}^{N} w_i r_i \right\rangle = \left\langle \boldsymbol{w}^T\boldsymbol{r}\right\rangle_c = \boldsymbol{w}^T\boldsymbol{\mu}^c \\ \sigma_{N,c}^2 &= \left\langle \left(\sum_{i=1}^{N} w_i r_i\right)^2 \right\rangle - \left\langle \sum_{i=1}^{N} w_i r_i \right\rangle^2 = \boldsymbol{w}^T\boldsymbol{\Gamma}_c\boldsymbol{w}\end{aligned} \quad \text{(A.36)}$$

The matrix $\boldsymbol{\Gamma}$ is the arithmetic mean of the covariance matrices computed with trials of fixed each choice, $\Gamma = \left(\boldsymbol{\Gamma}_A + \boldsymbol{\Gamma}_B\right)/2$, and $\boldsymbol{\Gamma}_{c,ij} = \langle r_i r_j\rangle_c - \langle r_i\rangle_c\langle r_j\rangle_c$

The set of weights $\boldsymbol{w}$ that maximizes $\mathrm{CP_N}$ can be found by writing Δ_N^2,

$$\Delta_N^2 = \frac{\boldsymbol{w}^T\left(\boldsymbol{\mu}^A - \boldsymbol{\mu}^B\right)\left(\boldsymbol{\mu}^A - \boldsymbol{\mu}^B\right)^T\boldsymbol{w}}{\boldsymbol{w}^T\boldsymbol{\Gamma}\boldsymbol{w}} \quad \text{(A.37)}$$

This quantity is the Fisher's Linear Discriminant between classes A and B. It measures the ratio between the squared difference of the mean values of rN in each class and the average variance within each decision. Assuming that Γ has full rank, the optimal vector $\boldsymbol{w}$ is,

$$\boldsymbol{w}^{opt} = \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{\mu}^A - \boldsymbol{\mu}^B\right) \quad \text{(A.38)}$$

and the optimal $\mathrm{CP_N}$ is,

$$\mathrm{CP_N} = \frac{1}{2}\mathrm{erfc}\left(-\frac{\Delta_N}{2}\right), \quad \Delta_N^2 = \sqrt{(\boldsymbol{\mu}^A - \boldsymbol{\mu}^B)\,\boldsymbol{\Gamma}^{-1T}\,(\boldsymbol{\mu}^A - \boldsymbol{\mu}^B)} \tag{A.39}$$

Notice that solving this optimization problem requires knowledge of the entire covariance matrix, Γ, an information that is experimentally very difficult to obtain because it involves the simultaneous recording of the entire population of neurons. In the particular case in which neurons belong to two discrete homogeneous pools, the $\mathrm{CP_N}$ reduces to the expression given in A.31-A.33.

A.6 Choice probability and Correlation Structure

The correlation coefficient R_{ij} between a pair of neurons (i, j) computed over trials of both decisions A and B is defined as,

$$R_{ij} = \frac{\mathrm{cov}(r_i, r_j)}{\sigma_i \sigma_j} \tag{A.40}$$

$$\begin{aligned} \mathrm{cov}(r_i, r_j) &= \langle (r_i - \mu_i)\,(r_j - \mu_j) \rangle \\ \sigma_i &= \langle (r_i - \mu_i) \rangle \\ \sigma_j &= \langle (r_j - \mu_j) \rangle \end{aligned} \tag{A.41}$$

where r_x is firing rate of neuron $x = i, j$ and μ_x is its mean value over all trials. The covariance $\mathrm{cov}(r_i, r_j)$ can be rewritten segregating trials according to the animal's decision,

$$\mathrm{cov}(r_i, r_j) = \frac{N_A}{N}\langle (r_i - \mu_i)\,(r_j - \mu_j) \rangle_A + \frac{N_B}{N}\langle (r_i - \mu_i)\,(r_j - \mu_j) \rangle_B \tag{A.42}$$

where N_A and N_B are the number of trials of each type and $N = N_A + N_B$. The mean firing rate over all trials can be expressed for both neurons ($x = i, j$) as,

$$\begin{aligned} \mu_x &= \frac{N_A}{N}\mu_x^A + \frac{N_B}{N}\mu_x^B \\ &= \mu_x^A - \frac{N_B}{N}\Delta\mu_x \\ &= \mu_x^B + \frac{N_A}{N}\Delta\mu_x \end{aligned} \tag{A.43}$$

with

$$\begin{aligned} \Delta\mu_x &= \mu_x^A - \mu_x^B \\ \mu_x^A &= \langle r_x \rangle_A \\ \mu_x^B &= \langle r_x \rangle_B \end{aligned} \tag{A.44}$$

Then, the covariance is,

$$\begin{aligned} \mathrm{cov}(r_i, r_j) = &\frac{N_A}{N}\langle \left(r_i - \mu_i^A\right)\left(r_j - \mu_j^A\right)\rangle_A + \\ &+ \frac{N_B}{N}\langle \left(r_i - \mu_i^B\right)\left(r_j - \mu_j^B\right)\rangle_B + \\ &+ \frac{N_A N_B^2 + N_B N_A^2}{N^3}\Delta\mu_i\Delta\mu_j \end{aligned} \tag{A.45}$$

which can be expressed as,

$$\mathrm{cov}(r_i, r_j) = \frac{N_A}{N}\gamma_{i,j}^A + \frac{N_B}{N}\gamma_{i,j}^B + \frac{N_A N_B^2 + N_B N_A^2}{N^3}\Delta\mu_i\Delta\mu_j \tag{A.46}$$

where $\gamma_{i,j}^A$ y $\gamma_{i,j}^B$ are the covariances computed segregating trials according to types A and B, respectively. A similar decomposition can be done for the variances of the firing rates ($x = i, j$),

$$\sigma_x^2 = \frac{N_A}{N}\sigma_{x,A}^2 + \frac{N_B}{N}\sigma_{x,A}^2 + \frac{N_A N_B^2 + N_B N_A^2}{N^3}\Delta\mu_x^2 \tag{A.47}$$

Considering Equations (A.46) and (A.47) together,

$$R_{i,j} = \frac{\frac{N_A}{N}\sigma_{i,A}\sigma_{j,A}\rho_{i,j}^A + \frac{N_B}{N}\sigma_{i,B}\sigma_{j,B}\rho_{i,j}^B + K\Delta\mu_i\Delta\mu_j}{\sqrt{\frac{N_A}{N}\sigma_{i,A}^2 + \frac{N_B}{N}\sigma_{i,B}^2 + K\Delta\mu_i^2}\sqrt{\frac{N_A}{N}\sigma_{j,A}^2 + \frac{N_B}{N}\sigma_{j,B}^2 + K\Delta\mu_j^2}} \tag{A.48}$$

where $K = \frac{N_A N_B^2 + N_B N_A^2}{N^3}$.

For the particular case in which the variances of the two neurons are equal in both type of trials and the number of A and B trials is the same,

$$R_{ij} = \frac{\frac{1}{2}\left(\rho_{i,j}^A + \rho_{i,j}^B\right) + \frac{1}{4}\delta_i\delta_j}{\sqrt{\left[1 + \left(\frac{\delta_i}{2}\right)^2\right]\left[1 + \left(\frac{\delta_j}{2}\right)^2\right]}} \tag{A.49}$$

that is Equation 3.7 of the main text. Equation (A.49) expresses the usual correlation coefficient R in terms of correlations coefficients computed segregating trials according to the animal's choice (ρ_{ij}^A and ρ_{ij}^B) and in terms of δ_i and δ_j (which is the quantity that determines the CP index, Equation A.10).

To obtain further insight about the meaning of this expression and to justify the approximate expression for the CP index (Equation 3.8), let us consider the Taylor expansion of $\delta_{ij} = (\delta_i + \delta_j)/2$ around $\epsilon = (\delta_i - \delta_j)/2 \sim 0$. Under this approximation,

$$\delta_i = \delta_{ij}^0 + \frac{\epsilon^2}{2} f\left(R_{ij}, \rho_{ij}^{AB}\right) + O(\epsilon^4) \tag{A.50}$$

with

$$\left(\frac{\delta_0}{2}\right)^2 = \frac{R_{ij} - \rho_{ij}}{1 - R_{ij}} \tag{A.51}$$

$$f\left(R_{ij}, \rho_{ij}^{AB}\right) = \frac{-1 - 3(\delta_{ij}^0)^2 + \rho_{ij}^{AB}\left[(\delta_{ij}^0)^2 - 1\right]}{4(\rho_{ij}^{AB} - 1)\left[\delta_{ij}^0 + (\delta_{ij}^0)^3\right]} \tag{A.52}$$

where $\rho_{12}^{AB} = (\rho_{12}^A + \rho_{12}^B)/2$. Comparing the first and second term of the expansion we conclude that the accuracy of this approximation is good if $R_{ij} - \rho_{ij}^{AB} \neq 0$. For example, when $R_{ij} - \rho_{ij}^{AB} > 0.15$, a pair (i, j) with δ_i and δ_j differing by a 30% ($\epsilon = 0.3$) will give a relative error of less than 9%.

When Equation A.49 is applied to a pair of neurons (1,2) within the same pool we have,

$$\left(\frac{\delta_0}{2}\right)^2 = \frac{R_{12}^w - \rho_{12}^w}{1 - R_{12}^w} \tag{A.53}$$

where $\delta_0 = (\delta_1 + \delta_2)/2$. This is Equation 3.8 in the main text. Notice from Equation A.53 that,

$$R_{12}^w - \rho_{12}^{w,AB} \geq 0 \tag{A.54}$$

which means that correlations for neurons in the same pool, computed with fixed-choice trials, are smaller than those obtained with the whole set of trials.

For small δ_0 the CP index of neurons in the pair (1,2) can be linearized, obtaining a rather simple relationship between this index and noise correlations,

$$\left(\mathrm{CP} - \frac{1}{2}\right)^2 \sim \frac{1}{\pi}\frac{R_{12}^w - \rho_{12}^w}{1 - R_{12}^w} \tag{A.55}$$

and the corresponding average over the population of pairs. The error introduced by this linearization becomes significant at large values of CP (8% for CP = 0.75).

Instead, if we consider that a pair of neurons (1,3) belonging to different pools have opposite sign of the difference of mean responses in the two conditions (that is, $\delta_1\delta_3 < 0$), from Equation A.49 we have that,

$$\begin{aligned} R_{13}^b &= \frac{\rho_{13}^{b,A} + \rho_{13}^{b,B} + \frac{1}{2}\delta_1\delta_3}{2\sqrt{\left[1 + \left(\frac{\delta_1}{2}\right)^2\right]\left[1 + \left(\frac{\delta_3}{2}\right)^2\right]}} \\ R_{13}^b &\leq \rho_{13}^{b,A} + \rho_{13}^{b,B} + \frac{1}{2}\delta_1\delta_3 \\ R_{13}^b &\leq \rho_{13}^{b,AB} \end{aligned} \tag{A.56}$$

If for a pair of neurons from different pools $\delta_3 \sim -\delta_1$, then $\delta_0 = \frac{1}{2}\left(|\delta_1| + |\delta_3|\right)$ can be approximated in terms of the correlation coefficients of the pair (1,3) as,

$$\left(\frac{\delta_0}{2}\right)^2 \sim \frac{\rho_{13}^b - R_{13}^b}{1 - R_{13}^b}, \tag{A.57}$$

where we defined $\rho_{13}^b = (\rho_{13}^{b,A} + \rho_{13}^{b,B})/2$.

As a more complex example, we now consider a two-pool network satisfying both conditions $\delta_1 \sim \delta_2 \sim -\delta_3$. Equations A.53 and A.57 can be seen as a constraint on the correlation structure of the network. Using this constraint, and replacing pair-wise correlations by their population-averaged values ($\bar{R}^w$,$\bar{R}^b$,$\bar{\rho}^w$,$\bar{\rho}^b$)

we obtain

$$\overline{\mathrm{CP}} \sim \frac{1}{2}\mathrm{erfc}\left(-\sqrt{\frac{(R^w - R^b) - (\rho^{w,AB} - \rho^{b,AB})}{2 - (R^w - R^b)}}\right). \tag{A.58}$$

which is Equation 3.9 of the main text.

A similar study can be done for the other generalized indices. For a pair of neurons (1,2) within the same neural pool and a pair of neurons (1,3) between different pools, we have,

$$\begin{aligned} \Delta_{2,w} &= \frac{\sqrt{2}\delta_0}{\sqrt{1+\rho_{12}^w}}, \quad \left(\frac{\delta_0}{2}\right)^2 \sim \frac{R_{12}^w - \rho_{12}^w}{1 - R_{12}^w} \\ \Delta_{2,b} &= \frac{(1-D)\,\delta_0}{\sqrt{1 + D^2 + D\rho_{13}^b}}, \quad \left(\frac{\delta_0}{2}\right)^2 \sim \frac{\rho_{13}^b - R_{13}^b}{1 - R_{13}^b} \end{aligned} \tag{A.59}$$

The $\mathrm{CP_N}$ index can be related to the choice-conditioned correlation coefficients as,

$$\Delta_{\mathrm{N}} = \frac{\sqrt{N}\left(\bar{\delta}_+ D \bar{\delta}_-\right)}{\sqrt{(1+D^2)\left[1 + (N-1)\,\bar{\rho}^w\right] + 2DN\bar{\rho}^b}} \tag{A.60}$$

where $\bar{\delta}_+$ and $\bar{\delta}_-$ can be obtained averaging the expression in Equation A.51 over the population of positive and negative neuronal pairs, respectively.

A.7 Dependence of the CP_N index on the number of neurons

In order to explore the dependence of the CP_N index on the number of neurons, we calculated this index taking from our database subsets of neurons with different sizes. We performed this analysis on the population of delay-activity PM neurons. We randomly selected the same number of neurons from the positive and negative pools and averaged the results over 100 repetitions for each population size.

The CP_N index for a two pools system with N neurons each can be expressed as,

$$\Delta_N = \frac{\sqrt{N}\left[\left(\bar{\mu}_+^A - \bar{\mu}_+^B\right) + D\left(\bar{\mu}_-^A - \bar{\mu}_-^B\right)\right]}{\sqrt{\bar{\sigma}_+^2 + (N-1)\,\bar{\gamma}_{++}^c + D^2\left(\sigma_-^2 + (N-1)\,\bar{\gamma}_{--}^c\right) + 2DN\bar{\gamma}_{+-}^c}} \tag{A.61}$$

Figure A.4 (top row) shows the time course of CP_N, the numerator and the denominator of Equation A.61, for N taking values between 5 and 80 per pool. For small N the CPN index increases with N while at larger values of N it saturates. Figure A.4 (bottom row) shows the dependence on N of the CP_N index, num^2 and den^2 for a fixed time, t=1s.

The dependence of CP_N on N can be explained in terms of Equation A.61. For small N, $\bar{\sigma}_+^2 >> (N-1)\,\bar{\gamma}_{++}^c$ and $\sigma_-^2 >> (N-1)\,\bar{\gamma}_{--}^c$. Therefore the CP_N index increases with N following the dependence on N of the numerator. For large N, $\bar{\sigma}_+^2 << (N-1)\,\bar{\gamma}_{++}^c$ and $\sigma_-^2 << (N-1)\,\bar{\gamma}_{--}^c$, therefore Equation A.61 can be approximated as,

$$\Delta_N = \frac{\sqrt{N}\left[\left(\bar{\mu}_+^A - \bar{\mu}_+^B\right) + D\left(\bar{\mu}_-^A - \bar{\mu}_-^B\right)\right]}{\sqrt{N}\sqrt{\bar{\gamma}_{++}^c + D^2\bar{\gamma}_{--}^c + 2D\bar{\gamma}_{+-}^c}} \tag{A.62}$$

which explains the linear dependence of num^2 and den^2 on N (Figure A.4, bottom row) and the saturation of CP_N at large N.

Appendix B

Supplemental Material to Chapter 4

B.1 Examples of realigned activity of FA and CR trials

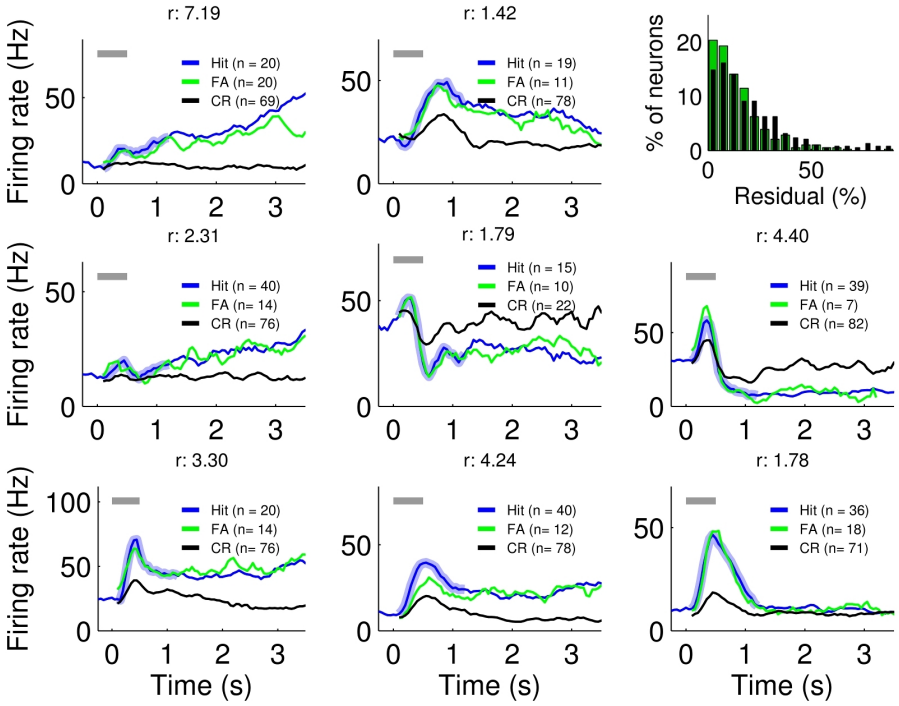

Figure B.1: **Examples of realigned activity of FA and CR trials** Examples of realigned averaged activity of FA (green) and CR (black) trials for 9 neurons with different firing temporal profiles (analogous to Figure 4.2c). Single trials in which a FA event was detected were realigned according to the detected time. Blue trace corresponds to the average over strong amplitude Hit trials and the shadow indicates the segment used as template. The number of trials for each condition is indicated by n. The top-right panel shows the histogram over neurons of the difference between realigned FA and Hit profiles (compared with the same measure for CR's realigned profile). The quantity, denoted r, was calculated as the sum of squared residuals, measured as a percent of the mean Hit activity. The distribution of this measure over the recorded neurons is shown for FA (green) and CR (black). Lower residuals are found for the realigned FA profiles than for the realigned CR profile, indicating a significant better match of the former. Indeed, 30% of the recorded neurons had residuals lower than 7% (for visual reference, the values for the example neurons is indicated in each panel).

B.2 Neural trajectories obtained separately for the two subjects

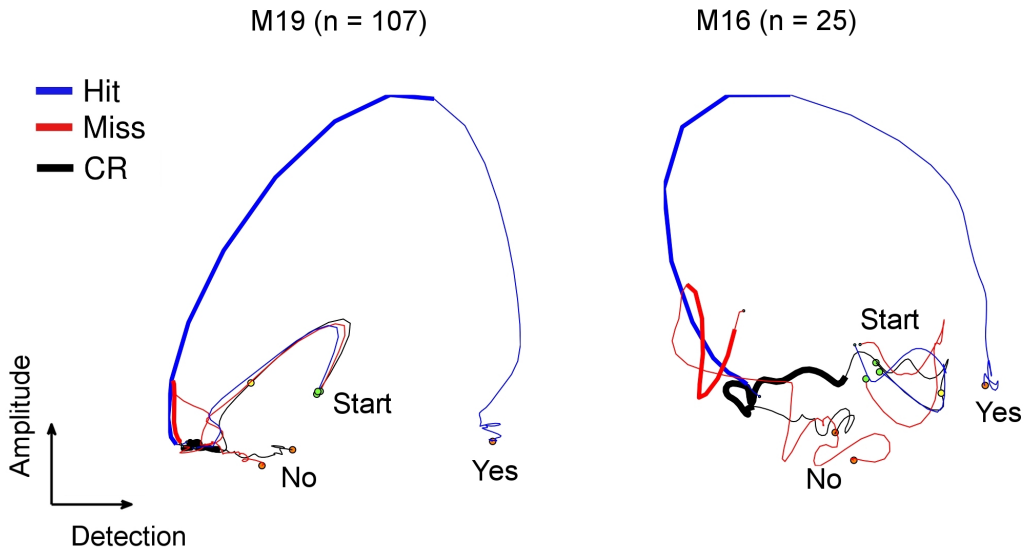

Figure B.2: **Neural trajectories obtained separately for the two subjects** Figure 4.4 in the main text was obtained by combining data from two monkeys (M16 and M19). Here we present the same analysis performed separately for each subject. Although the smaller number of neurons deteriorates the clarity of the neural trajectories -specially in subject M16- the same conclusions as in the main text can be reached. The average neural trajectories during Hit (blue), Miss (red) and CR (black) trials were projected onto two task-related axes (stimulus amplitude and stimulus detection). As in Figure 4.4 of the main text, the trajectories are plotted from the beginning of the trial (green circles) to end of the delay period (orange circles). Stimulus-present conditions are plotted until 1.5 s and realigned at the stimulus onset time. Thick blue and red traces indicate the period of stimulation. The thick black line denotes the possible stimulation window (1.5 s to 3.5s). Units are arbitrary.

B.3 Neural trajectories projected onto the the principal components

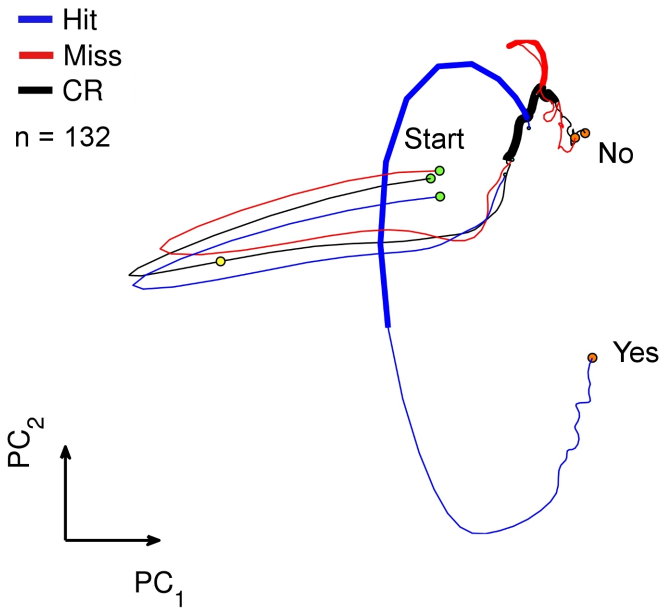

Figure B.3: **Neural trajectories projected onto the the principal components** Average neural trajectories during Hit (blue), Miss (red) and CR (black) trials projected onto the first two principal components of the data. The modulation in CR's trajectory during the window of possible stimulation (1.5-3.5 s, thick black line) is also visible here, and does not depend crucially on the selected axes (Figure 4.4, main text). The trajectories are plotted from the beginning of the trial (green circles) to end of the delay period (orange circles). Stimulus-present conditions are plotted until 1.5 s and realigned at the stimulus onset time. Thick blue and red traces indicate the period of stimulation. (1.5 s to 3.5s). Units are arbitrary.

B.4 Neural trajectories close to the separatrix

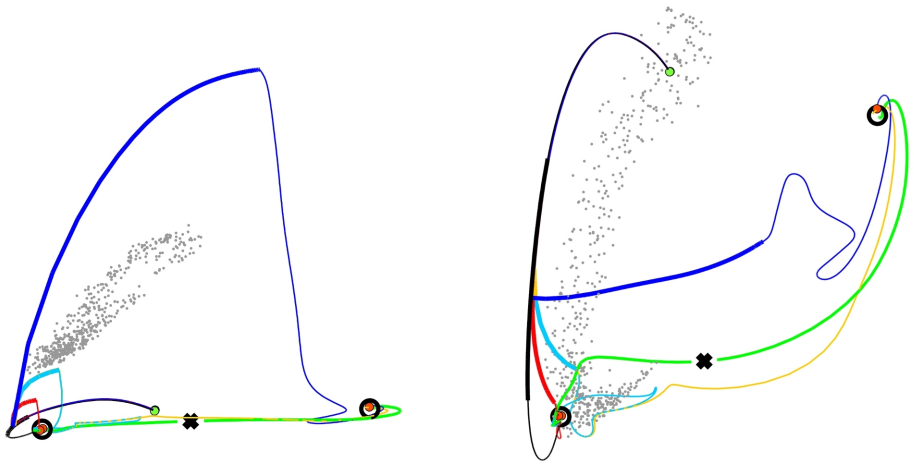

Figure B.4: **Neural trajectories close to the separatrix** Same simulations as in Figure 4.7 but including two trajectories obtained by stimulating with 'borderline' amplitudes (cyan and yellow). (*Left*) Same projection as in Figure 4.7. (*Right*) A rotation in neural space to visualize how the two 'borderline' trajectories travel close to the separatrix and approach the saddle point (black cross). Afterward, each of them travels to a different attractor.

Nomenclature

CR	correct rejection
FA	false alarm
PMc	premotor cortex
$S1$	primary somatosensory cortex
$S2$	secondary somatosensory cortex
STD	signal detection theory

List of Figures

LIST OF FIGURES

References

L. Abbott and P. Dayan. The effect of correlated variability on the accuracy of a population code. *Neural computation*, 11(1):91–101, 1999. 3

G. Aston-Jones and J. D. Cohen. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, 28:403–450, 2005. 21

B. B. Averbeck and D. Lee. Neural noise and movement-related codes in the macaque supplementary motor area. *The Journal of neuroscience*, 23(20):7630–7641, 2003. 16

B. B. Averbeck, P. E. Latham, and A. Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–366, 2006. 3, 83, 87

J. M. Beck, W. J. Ma, R. Kiani, T. Hanks, A. K. Churchland, J. Roitman, M. N. Shadlen, P. E. Latham, and A. Pouget. Probabilistic population codes for bayesian decision making. *Neuron*, 60(6):1142–1152, 2008. 22

K. H. Britten, M. N. Shadlen, W. T. Newsome, and J. A. Movshon. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *The Journal of Neuroscience*, 12(12):4745–4765, 1992. 3

K. H. Britten, W. T. Newsome, M. N. Shadlen, S. Celebrini, and J. A. Movshon. A relationship between behavioral choice and the visual responses of neurons in macaque mt. *Visual neuroscience*, 13(01):87–100, 1996. 25, 28

C. D. Brody. Correlations without synchrony. *Neural computation*, 11(7):1537–1551, 1999. 34

C. D. Brody, A. Hernández, A. Zainos, and R. Romo. Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cerebral cortex*, 13(11):1196–1207, 2003. 50

D. V. Buonomano and W. Maass. State-dependent computations: spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, 10(2):113–125, 2009. 73

F. Carnevale, V. de Lafuente, R. Romo, and N. Parga. Internal signal correlates neural populations and biases perceptual decision reports. *Proceedings of the National Academy of Sciences*, 109(46):18938–18943, 2012. 26, 42, 49, 59

F. Carnevale, V. de Lafuente, R. Romo, and N. Parga. An optimal decision population code that accounts for correlated variability unambiguously predicts a subjects choice. *Neuron*, 80(6):1532–1543, 2013. 59, 63

D. J. Chalmers. Facing up to the problem of consciousness. *Journal of consciousness studies*, 2(3):200–219, 1995. 3

M. R. Cohen and J. H. Maunsell. A neuronal population measure of attention predicts behavioral performance on individual trials. *The Journal of Neuroscience*, 30(45):15241–15253, 2010. 51

M. R. Cohen and J. H. Maunsell. Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron*, 70(6):1192–1204, 2011. 51

M. R. Cohen and W. T. Newsome. Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *The Journal of Neuroscience*, 29(20):6635–6648, 2009. 26, 83, 87

C. Constantinidis and P. S. Goldman-Rakic. Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *Journal of neurophysiology*, 88(6):3487–3497, 2002. 17

E. P. Cook and J. H. Maunsell. Dynamics of neuronal responses in macaque mt and vip during motion detection. *Nature neuroscience*, 5(10):985–994, 2002. 13, 22

Á. Correa, J. Lupiáñez, and P. Tudela. Attentional preparation based on temporal expectancy modulates processing at the perceptual level. *Psychonomic bulletin & review*, 12(2):328–334, 2005. 56, 75

J. Coull and A. Nobre. Dissociating explicit timing from temporal expectation with fmri. *Current opinion in neurobiology*, 18(2):137–144, 2008. 56

J. De La Rocha, B. Doiron, E. Shea-Brown, K. Josić, and A. Reyes. Correlation between neural spike trains increases with firing rate. *Nature*, 448(7155):802–806, 2007. 83, 87

V. de Lafuente and R. Romo. Neuronal correlates of subjective sensory experience. *Nature neuroscience*, 8(12):1698–1703, 2005. 8, 13, 14, 16, 22, 26, 27, 36, 52, 55, 56, 59, 63, 68, 76, 77

V. de Lafuente and R. Romo. Neural correlate of subjective sensory experience gradually builds up across cortical areas. *Proceedings of the National Academy of Sciences*, 103(39):14266–14271, 2006. 8, 13, 14, 16, 22, 25, 26, 27, 28, 36, 44, 52, 55, 56, 59, 64, 76, 77

V. de Lafuente and R. Romo. Dopamine neurons code subjective sensory experience and uncertainty of perceptual decisions. *Proceedings of the National Academy of Sciences*, 108(49):19767–19771, 2011. 21

K. Doya. Modulators of decision making. *Nature neuroscience*, 11(4):410–416, 2008. 21

A. S. Ecker, P. Berens, G. A. Keliris, M. Bethge, N. K. Logothetis, and A. S. Tolias. Decorrelated neuronal firing in cortical microcircuits. *Science*, 327 (5965):584–587, 2010. 22, 34, 48

L. Fenno, O. Yizhar, and K. Deisseroth. The development and application of optogenetics. *Annual review of neuroscience*, 34:389–412, 2011. 85, 89

B. U. Forstmann, S. Brown, G. Dutilh, J. Neumann, and E.-J. Wagenmakers. The neural substrate of prior information in perceptual decision making: a model-based analysis. *Frontiers in human neuroscience*, 4, 2010. 55, 74

T. J. Gawne and B. J. Richmond. How independent are the messages carried by adjacent inferior temporal cortical neurons? *The Journal of neuroscience*, 13 (7):2758–2771, 1993. 26

N. S. Ghandehari. Attention modulates synchronized neuronal firing in primate somatosensory cortex. page 2012, 2000. 73

G. M. Ghose and D. W. Bearl. Attention directed by expectations enhances receptive fields in cortical area mt. *Vision research*, 50(4):441–451, 2010. 56, 75

G. M. Ghose and J. H. Maunsell. Attentional modulation in visual cortex depends on task timing. *Nature*, 419(6907):616–620, 2002. 56, 75

C. D. Gilbert and M. Sigman. Brain states: top-down influences in sensory processing. *Neuron*, 54(5):677–696, 2007. 4, 55

J. I. Gold and M. N. Shadlen. Representation of a perceptual decision in developing oculomotor commands. *Nature*, 404(6776):390–394, 2000. 13, 14, 21

J. I. Gold and M. N. Shadlen. Neural computations that underlie decisions about sensory stimuli. *Trends in cognitive sciences*, 5(1):10–16, 2001. 4, 25, 45

J. I. Gold and M. N. Shadlen. The neural basis of decision making. *Annu. Rev. Neurosci.*, 30:535–574, 2007. 2, 84, 88

P. Goldman-Rakic. Cellular basis of working memory. *Neuron*, 14(3):477–485, 1995. 21

D. M. Green, J. A. Swets, et al. *Signal detection theory and psychophysics*, volume 1. Wiley New York, 1966. 5, 25, 28, 54, 73

R. M. Haefner, S. Gerwinn, J. H. Macke, and M. Bethge. Inferring decoding strategies from choice probabilities in the presence of correlated variability. *Nature neuroscience*, 16(2):235–242, 2013. 26, 83, 87

D. P. Hanes and J. D. Schall. Neural control of voluntary movement initiation. *Science*, 274(5286):427–430, 1996. 13, 14, 21

T. D. Hanks, M. E. Mazurek, R. Kiani, E. Hopp, and M. N. Shadlen. Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *The Journal of Neuroscience*, 31(17):6339–6352, 2011. 55, 74, 75

H. Heekeren, S. Marrett, P. Bandettini, and L. Ungerleider. A general mechanism for perceptual decision-making in the human brain. *Nature*, 431(7010):859–862, 2004. 28, 45

A. Hernández, A. Zainos, and R. Romo. Temporal evolution of a decision-making process in medial premotor cortex. *Neuron*, 33(6):959–972, 2002. 13, 21

A. Hernández, V. Nácher, R. Luna, M. Alvarez, A. Zainos, S. Cordero, L. Camarillo, Y. Vázquez, L. Lemus, and R. Romo. Procedure for recording the simultaneous activity of single neurons distributed across cortical areas during sensory discrimination. *Proceedings of the National Academy of Sciences*, 105 (43):16785–16790, 2008. 51

A. Hernández, V. Nácher, R. Luna, A. Zainos, L. Lemus, M. Alvarez, Y. Vázquez, L. Camarillo, and R. Romo. Decoding a perceptual decision process across cortex. *Neuron*, 66(2):300–314, 2010. 4, 13, 14, 21, 25, 28, 49, 50

G. M. Hoerzer, R. Legenstein, and W. Maass. Emergence of complex computational structures from chaotic neural networks through reward-modulated hebbian learning. *Cerebral Cortex*, 24(3):677–690, 2014. 74

H. Jaeger and H. Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004. 73

P. Janssen and M. N. Shadlen. A representation of the hazard rate of elapsed time in macaque area lip. *Nature neuroscience*, 8(2):234–241, 2005. 56, 73

S. Jaramillo and A. M. Zador. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nature neuroscience*, 14(2):246–251, 2011. 56, 74, 75

J. K. Jun, P. Miller, A. Hernández, A. Zainos, L. Lemus, C. D. Brody, and R. Romo. Heterogenous population coding of a short-term memory and decision task. *The Journal of Neuroscience*, 30(3):916–929, 2010. 51

U. R. Karmarkar and D. V. Buonomano. Timing in the absence of clocks: encoding time in neural network states. *Neuron*, 53(3):427–438, 2007. 73

S. Katzner, S. Treue, and L. Busse. Improving behavioral performance under full attention by adjusting response criteria to changes in stimulus predictability. *Journal of vision*, 12(10):1, 2012. 56

J.-N. Kim and M. N. Shadlen. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature neuroscience*, 2(2):176–185, 1999. 13

R. Laje and D. V. Buonomano. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nature neuroscience*, 16(7):925–933, 2013. 74

L. Lemus, A. Hernández, R. Luna, A. Zainos, V. Nácher, and R. Romo. Neural correlates of a postponed decision report. *Proceedings of the National Academy of Sciences*, 104(43):17174–17179, 2007. 50

L. Lemus, A. Hernández, and R. Romo. Neural encoding of auditory discrimination in ventral premotor cortex. *Proceedings of the National Academy of Sciences*, 106(34):14640–14645, 2009. 50

M. I. Leon and M. N. Shadlen. Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron*, 38(2):317–327, 2003. 73

K. Louie and M. A. Wilson. Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, 29(1):145–156, 2001. 75

R. D. Luce. *Response Times: Their Role in Inferring Elementary Mental Organization3*. Number 8. Oxford University Press, 1986. 56

W. Maass, P. Joshi, and E. D. Sontag. Computational aspects of feedback in neural circuits. *PLOS Computational Biology*, 3(1):e165, 2007. 74

V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474): 78–84, 2013. 64

T. Masquelier. Neural variability, or lack thereof. *Frontiers in computational neuroscience*, 7, 2013. 3

K. Merten and A. Nieder. Active encoding of decisions about stimulus absence in primate prefrontal cortex neurons. *Proceedings of the National Academy of Sciences*, 109(16):6289–6294, 2012. 21

E. K. Miller. The prefontral cortex and cognitive control. *Nature reviews neuroscience*, 1(1):59–65, 2000. 14

V. B. Mountcastle, W. H. Talbot, I. Darian-Smith, and H. H. Kornhuber. Neural basis of the sense of flutter-vibration. *Science*, 155(3762):597–600, 1967. 7

V. B. Mountcastle, W. H. Talbot, H. Sakata, and J. Hyvarinen. Cortical neuronal mechanisms in flutter-vibration studied in unanesthetized monkeys. neuronal periodicity and frequency discrimination. *J Neurophysiol*, 32(3):452–484, 1969. 7

H. Nienborg and B. Cumming. Correlations between the activity of sensory neurons and behavior: how much do they tell us about a neuron's causality? *Current opinion in neurobiology*, 20(3):376–381, 2010. 26, 35, 83, 87

H. Nienborg and B. G. Cumming. Decision-related activity in sensory neurons reflects more than a neurons causal effect. *Nature*, 459(7243):89–92, 2009. 4, 14, 21

H. Nienborg, M. R. Cohen, and B. G. Cumming. Decision-related activity in sensory neurons: correlations among neurons and with behavior. *Annual review of neuroscience*, 35:463–483, 2012. 4, 26, 35, 83, 87

A. C. Nobre. Orienting attention to instants in time. *Neuropsychologia*, 39(12): 1317–1328, 2001. 5, 56

A. C. Nobre, A. Correa, and J. T. Coull. The hazards of time. *Current opinion in neurobiology*, 17(4):465–470, 2007. 5, 56

B. Pesaran, M. J. Nelson, and R. A. Andersen. Free choice activates a decision circuit between frontal and parietal cortex. *Nature*, 453(7193):406–409, 2008. 26, 28

M. L. Platt and P. W. Glimcher. Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741):233–238, 1999. 22

M. I. Posner. Orienting of attention. *Quarterly journal of experimental psychology*, 32(1):3–25, 1980. 5

R. P. Rao. Decision making under uncertainty: a neural model based on partially observable markov decision processes. *Frontiers in computational neuroscience*, 4, 2010. 22

V. Rao, G. C. DeAngelis, and L. H. Snyder. Neural correlates of prior expectations of motion in the lateral intraparietal and middle temporal areas. *The Journal of Neuroscience*, 32(29):10063–10074, 2012. 55, 74

R. Ratcliff and G. McKoon. The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, 20(4):873–922, 2008. 55, 74

A. Renart and C. K. Machens. Variability in neural activity and behavior. *Current opinion in neurobiology*, 25:211–220, 2014. 3, 75

A. Renart, J. de la Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris. The asynchronous state in cortical circuits. *science*, 327(5965): 587–590, 2010. 22, 34, 48, 83, 87

G. Rohenkohl, A. M. Cravo, V. Wyart, and A. C. Nobre. Temporal expectation improves the quality of sensory information. *The Journal of Neuroscience*, 32 (24):8424–8428, 2012. 56, 75

R. Romo and V. de Lafuente. Conversion of sensory signals into perceptual decisions. *Progress in neurobiology*, 103:41–75, 2013. 2, 4, 25, 45, 55, 84, 88

R. Romo and E. Salinas. Touch and go: decision-making mechanisms in somatosensation. *Annual review of neuroscience*, 24(1):107–137, 2001. 2

R. Romo and E. Salinas. Flutter discrimination: neural codes, perception, memory and decision making. *Nature Reviews Neuroscience*, 4(3):203–218, 2003. 4, 25, 45

R. Romo and W. Schultz. Role of primate basal ganglia and frontal cortex in the internal generation of movements. iii. neuronal activity in the supplementary motor area. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, 91(3):396–407, 1991. 21

R. Romo, C. D. Brody, A. Hernández, and L. Lemus. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, 399(6735):470–473, 1999. 13, 21, 50

R. Romo, A. Hernández, A. Zainos, L. Lemus, and C. D. Brody. Neuronal correlates of decision-making in secondary somatosensory cortex. *Nature neuroscience*, 5(11):1217–1225, 2002. 13

R. Romo, A. Hernández, and A. Zainos. Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron*, 41(1):165–173, 2004. 13, 21

E. Salinas, A. Hernandez, A. Zainos, and R. Romo. Periodicity and firing rate as candidate neural codes for the frequency of vibrotactile stimuli. *The Journal of Neuroscience*, 20(14):5503–5515, 2000. 13

J. D. Schall. Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience*, 2(1):33–42, 2001. 21, 22

C. Scheibe, R. Schubert, W. Sommer, and H. R. Heekeren. Electrophysiological evidence for the effect of prior probability on response preparation. *Psychophysiology*, 46(4):758–770, 2009. 56

M. N. Shadlen and R. Kiani. Decision making as a window on cognition. *Neuron*, 80(3):791–806, 2013. 3

M. N. Shadlen and W. T. Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *The Journal of neuroscience*, 18(10):3870–3896, 1998. 3, 26

M. N. Shadlen and W. T. Newsome. Neural basis of a perceptual decision in the parietal cortex (area lip) of the rhesus monkey. *Journal of neurophysiology*, 86 (4):1916–1936, 2001. 13, 14, 21, 22

M. N. Shadlen, K. H. Britten, W. T. Newsome, and J. A. Movshon. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *The Journal of neuroscience*, 16(4):1486–1510, 1996. 13, 14, 21, 26, 34, 39, 42, 48

D. L. Sheinberg and N. K. Logothetis. The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences*, 94 (7):3408–3413, 1997. 22

K. V. Shenoy, M. Sahani, and M. M. Churchland. Cortical control of arm movements: a dynamical systems perspective. *Annual review of neuroscience*, 36: 337–359, 2013. 64

M. Siegel, A. K. Engel, and T. H. Donner. Cortical network dynamics of perceptual decision-making in the human brain. *Frontiers in human neuroscience*, 5: 21, 2011. 26, 28, 49

P. Simen, D. Contreras, C. Buck, P. Hu, P. Holmes, and J. D. Cohen. Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and Performance*, 35(6):1865, 2009. 55, 74

W. R. Softky and C. Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *The Journal of Neuroscience*, 13(1):334–350, 1993. 3

H. Sompolinsky, H. Yoon, K. Kang, and M. Shamir. Population coding in neuronal systems with correlated noise. *Physical Review E*, 64(5):051904, 2001. 3

P. N. Steinmetz, A. Roy, P. Fitzgerald, S. Hsiao, K. Johnson, and E. Niebur. Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, 404(6774):187–190, 2000. 75

M. G. Stokes, M. Kusunoki, N. Sigala, H. Nili, D. Gaffan, and J. Duncan. Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, 78(2):364–375, 2013. 64

C. Stosiek, O. Garaschuk, K. Holthoff, and A. Konnerth. In vivo two-photon calcium imaging of neuronal networks. *Proceedings of the National Academy of Sciences*, 100(12):7319–7324, 2003. 85, 89

C. Summerfield and E. Koechlin. A neural representation of prior information during perceptual inference. *Neuron*, 59(2):336–347, 2008. 55, 74

D. Sussillo and L. Abbott. Transferring learning from external to internal weights in echo-state networks with sparse connectivity. *PloS one*, 7(5):e37372, 2012. 68, 80

D. Sussillo and L. F. Abbott. Generating coherent patterns of activity from chaotic neural networks. *Neuron*, 63(4):544–557, 2009. 68, 80

D. Sussillo and O. Barak. Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural computation*, 25(3):626–649, 2013. 70, 81

J. Tanji. Sequential organization of multiple movements: involvement of cortical motor areas. *Annual review of neuroscience*, 24(1):631–651, 2001. 21

D. J. Tolhurst, J. Movshon, and A. Dean. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision research*, 23(8):775–785, 1983. 26

A. J. Yu and P. Dayan. Uncertainty, neuromodulation, and attention. *Neuron*, 46(4):681–692, 2005. 21

E. Zohary, M. N. Shadlen, and W. T. Newsome. Correlated neuronal discharge rate and its implications for psychophysical performance. 1994. 3, 26

Agradecimientos

Quisiera agradecer en primer lugar a mi director de tesis, N. Parga, por haberme dado esta oportunidad y por su constante confianza, ánimo y comprensión.

Debo un especial agradecimiento al Dr. Ranufo Romo, quien tan gentilmente me ha apoyado durante estos años. Ha sido un gran privilegio y un enorme placer cada visita a su laboratorio, especialmente por contar ahí con tantas y tan valiosas personas. De la misma forma, he disfrutado mucho de la colaboración con el Dr. Víctor de Lafuente, quien fue siempre tan amable, dispuesto y atento conmigo.

Agradezco al Dr. Omri Barak por recibirme en su laboratorio. Trabajar con él fue profesional y personalmente muy inspirador y su generosidad ha sido para mi, una bocanada de aire fresco.

Gracias especiales a los amigos y amigas que he tenido el placer de conocer en España. Les agradezco haber hecho de esta etapa un tiempo tan disfrutable.

Mi más profundo agradecimiento es hacia mi familia. Su amor y apoyo incondicional ha logrado hacerme sentir cerca estando lejos. Reservo el último y más importante lugar para Alina, quien me ha sabido entender cada día. Gracias por haber elegido compartir tu vida conmigo.