Temporal modulation of the dynamics of neuronal networks with cognitive function : experimental evidence and theoretical analysis

Laureline Logiaco

Thèse de Doctorat
de l'Université Pierre et Marie Curie

Spécialité : Cerveau, Cognition, Comportement

École doctorale ED3C

Présentée par
Laureline LOGIACO

Pour obtenir le grade de
Docteur
de l'Université Pierre et Marie Curie

Temporal modulation of the dynamics of neuronal networks with cognitive function: experimental evidence and theoretical analysis

---

Soutenue le 7 octobre 2015 devant le jury composé de :

Rapporteurs :	Stefano Fusi	-	Columbia University
	Jonathan Victor	-	Cornell University
Examinateurs :	Gianluigi Mongillo	-	Centre National de la Recherche Scientifique Université Paris-Descartes
	Emmanuel Procyk	-	Centre National de la Recherche Scientifique Université Claude Bernard
	Alfonso Renart	-	Champalimaud Center for the Unknown
Directeurs :	Angelo Arleo	-	Centre National de la Recherche Scientifique Université Pierre et Marie Curie
	Wulfram Gerstner	-	Ecole Polytechnique Fédérale de Lausanne
Président :	Philippe Faure	-	Centre National de la Recherche Scientifique Université Pierre et Marie Curie

Temporal modulation of the dynamics of neuronal networks with cognitive function: experimental evidence and theoretical analysis

Résumé vulgarisé

Nous nous sommes intéressés à la possible fonction de la structure temporelle de l'activité neuronale pendant l'évolution dynamique de réseaux de neurones qui peuvent sous-tendre des processus cognitifs.

Tout d'abord, nous avons caractérisé le code qui permet de lire l'information contenue dans des signaux neuronaux enregistrés dans le cortex cingulaire antérieur dorsal (CCAd). Le signal émis par les cellules neurales (les neurones) comporte une série de perturbations stéréotypiques du potentiel électrique, appelées potentiels d'action. Ce signal neuronal peut donc être caractérisé par le nombre et le temps des potentiels d'action émis durant une certaine situation comportementale.

Les données actuelles ont mis en évidence que le CCAd est impliqué dans les processus d'adaptation comportementale à de nouveaux contextes. Cependant, les mécanismes biologiques qui sous-tendent ces processus d'adaptation comportementale sont encore mal compris. Nos analyses suggèrent que la variabilité importante du nombre de potentiels d'action émis par les neurones, ainsi que la fiabilité temporelle conséquente de ces potentiels d'action (qui est améliorée par la présence de corrélations entre les temps d'émission des potentiels d'action), avantagent les réseaux neuronaux qui sont considérablement sensibles à la structure temporelle des signaux qu'ils reçoivent. Cet avantage se traduit par une augmentation de l'efficacité du décodage de signaux émis par des neurones du CCAd lorsque les singes changent de stratégie comportementale. Nous avons aussi cherché à déterminer les caractéristiques de la variabilité neuronale qui peuvent prédire la variabilité comportementale de l'animal. Quand nous avons séparé les données entre un groupe avec un grand nombre de potentiels d'action, et un groupe avec un faible nombre de potentiels d'action, nous n'avons pas trouvé pas de différence robuste et cohérente du comportement des animaux entre ces deux groupes. Par contre, nous avons trouvé que lorsque l'activité d'un neurone devient moins semblable à la réponse typique de ce neurone, les singes semblent répondre plus lentement pendant la tâche comportementale. Plus précisément, nous avons observé que l'activité d'un neurone semble pouvoir se différencier de sa réponse typique tantôt par une augmentation du nombre de potentiels d'actions émis, tantôt par une réduction de ce nombre. De plus, des imprécisions sur le temps d'émission des potentiels

d'action peuvent mener à une déviation du signal neuronal par rapport à la réponse typique.

Nos résultats suggèrent que le réseau, ou les réseaux de neurones qui reçoivent et décodent les signaux d'adaptation comportementale émis par le CCAd pourraient être adaptés à la détection de motifs définis à la fois dans l'espace (par l'identité du neurone ayant émis un potentiel d'action) et dans le temps (par le moment précis d'émission d'un potentiel d'action). Par conséquent, ces réseaux de neurones ne se comportent probablement pas comme des intégrateurs, qui sont des circuits dont le niveau d'activité reflète approximativement la somme des potentiels d'actions reçus pendant une certaine période de temps.

Dans un second temps, nous avons travaillé à mieux comprendre les méchanismes par lesquels le réseau de neurones décodant les signaux du CCAd pourait détecter un motif spatiotemporel. Pour cela, nous avons développé des équations qui réduisent la complexité du réseau en représentant l'ensemble des neurones par quelques statistiques représentatives. Nous avons choisi un modèle de neurone qui est capable de reproduire l'activité de neurones corticaux en réponse à des injections dynamiques de courant. Nous avons pu approximer la réponse de populations de neurones connectées de manière récurrente, lorsque les neurones émettent des potentiels d'action de façon assez irrégulière et asynchrone (ces caractéristiques sont communes dans les réseaux biologiques). Ce travail constitue une avancée méthodologique qui pourrait être le point de départ d'une étude des mécanismes par lesquels les reseaux de neurones récurrents, qui semblent être à l'origine des processus cognitifs, peuvent être influencés par la dynamique temporelle de leurs signaux d'entrée.

Abstract

We investigated the putative function of the fine temporal dynamics of neuronal networks for implementing cognitive processes.

First, we characterized the coding properties of spike trains recorded from the dorsal Anterior Cingulate Cortex (dACC) of monkeys. dACC is thought to trigger behavioral adaptation. We found evidence for (i) high spike count variability and (ii) temporal reliability (favored by temporal correlations) which respectively hindered and favored information transmission when monkeys were cued to switch the behavioral strategy. Also, we investigated the nature of the neuronal variability that was predictive of behavioral variability. High vs. low firing rates were not robustly associated with different behavioral responses, while deviations from a neuron-specific prototypical spike train predicted slower responses of the monkeys. These deviations could be due to increased or decreased spike count, as well as to jitters in spike times. Our results support the hypothesis of a complex spatiotemporal coding of behavioral adaptation by dACC, and suggest that dACC signals are unlikely to be decoded by a neural integrator.

Second, we further investigated the impact of dACC temporal signals on the downstream decoder by developing mean-field equations to analyze network dynamics. We used an adapting single neuron model that mimics the response of cortical neurons to realistic dynamic synaptic-like currents. We approximated the time-dependent population rate for recurrent networks in an asynchronous irregular state. This constitutes an important step towards a theoretical study of the effect of temporal drives on networks which could mediate cognitive functions.

Acknowledgements

This dissertation reflects in first place the influence of the many people who, throughout my life, have shaped and nurtured my mind. Among them were first my parents and my grand-parents, but also the many teachers who have been giving me the tools to think, and to express my thoughts. I would like to take the chance to thank them all, and to acknowledge the whole system by which my education could take place.

I would also like to thank the people who have encouraged me to work in research. In particular, I acknowledge the researchers who hosted a very young, very inexperienced and very useless first-year high school student in the Laboratoire de Bioénergétique Fondamentale et Appliquée of Joseph Fourier University. Their sincere interest in their work really motivated me during my studies, and they made me choose this very intriguing and weird job that research was for me at that time. In addition, I also want to stress that the internships tightly supervised by Pedro Gonçalves, Christian Machens and Sophie Denève have deeply shaped my path in research. The atmosphere and the extreme diversity and open-mindedness of the Group for Neural Theory at Ecole Normale Supérieure really defined the type of research that I aim to develop. I also have a thought for the people who helped me during my other internships, and more particularly for Ed Smith and Scott Livingston who made a lot of efforts to understand me, and to exchange thoughts my first study of neuronal data and animal behavior. Finally, I would like to thank Angelo Arleo for supervising me during the internship of my second year of master, which shaped the path of the doctoral work.

Concerning the doctoral studies, I would first like to thank my advisors for hosting me in their laboratories, and Emmanuel Procyk for providing me with an amazing data set that nurtured my whole doctoral studies. For the data analysis project, I have received invaluable help from Luca Leonardo Bologna (the master of computers), Jérémie Pinoteau, Eléonore Duvelle (who, in particular, reviewed a large part of this manuscript), Marie Rothé, Sergio Solinas, Dane Corneil, Felipe Gerhard, Tim Vogels, David Karstner and Christian Pozzorini. For the theoretical part, I got inspired and wisely advised by Moritz Deger and Tilo Schwalger, who left happy memories of black board brainstorming. I am also very grateful to have received advice from Carl Van Vreeswijk; without him, I would still be

erring among infinitely recurrent integral equations. I should also thank Skander Mensi and Christian Pozzorini; they were my first instructors for the Generalized Linear Model approach, they kindly shared their data with me, and were very patient in answering my many questions. Finally, I am very grateful to Aditya Gilra, Dane Corneil and Alex Seeholzer, for reviewing this dissertation. More generally, I would like to thank all the members and interns of the Aging in Vision and Action (previously Adaptive Neurocomputation) Laboratory, and of the Laboratory of Computational Neuroscience, for helping me at various points, and more particularly for correcting my overall bad skills for presenting my results. Also, I would like to acknowledge my father, who was eager to discuss my work, and Ritwik Niyogi, who offered me my first occasion to give a talk (and a nice stay in London!).

I also have a particular thought for Claudia Clopath, who gave me hope at a moment when I felt that nothing would ever work. Without her, I would never have dared to apply for a post-doctoral position while I had no paper.

Finally, and very importantly, I am very indebted to my family and friends (many of those being also colleagues!). I am most often limited by my negative emotions. Without receiving kind attention, I would simply wither away. I will cite an – incomplete – list of people. First, my parents and my grandparents, who have always been on my side during the hardest events of my life. In particular, I often took solace through the numerous messages sent by my mother, and through singing with my father. I also want to thank my brother, with whom I have been sharing the same roof for a quarter of a century. I like to think (but he may not agree) that him and I have to face very similar issues: we must first invent something that we like within our perception of the current framework, and then hope that other people will share our excitement. Therefore, sharing thoughts and music with him has always been reassuring. I would also like to thank my very old friends, including Laura Blondé, Violaine Mazas, Pauline Bertholio, Gabriel Besson and Anne Perez, for their priceless continuous kindness and support. I have a particular thought for Luca Leonardo Bologna, Jérémie Pinoteau, Dane Corneil, Olivia Gozel, Friedemann Zenke, Alexander Seeholzer, Vasiliki Liakoni, Thomas Van Pottelberg, Aditya Gilra, Marco Lehmann and his wife, and Eléonore Duvelle for their emotional support at and outside work. Finally, I will thank the many musicians who continuously help me to get through my life by expressing and sharing so well their inner worlds.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

An invitation to study the sensitivity of recurrent neuronal networks implementing cognitive computations to temporal signals

---

In this dissertation, we examine the characteristics and the functional relevance of the temporal structure of neuronal signals, in the context of cognitive processing and of recurrent neuronal networks. In order to explain the interest of this work, we will start by giving a very brief general introduction about the biological implementation of brain computations in general, and of cognitive processes in particular. We then introduce some classical models which are used to help explaining cognitive processes (such as memory or decision-making). Finally, we motivate the topic of the doctoral work, and we give a road map for the dissertation.

1.1 Background: neurons, networks, brain areas and brain processing

The computations performed by the brain are thought to occur through the dynamics of connected populations of neurons [Gerstner et al. (2014)]. The neurons are indeed often considered as the basic units of neuronal processing. They are connected together over different spatial scales, ranging from connections within a layer of a small patch of cortex [Avermann et al. (2012)] to connections between brain areas that implement different types of brain processing [Medalla and Barbas (2009); Boucsein et al. (2011)]. We first review basic single neuron properties, before sketching examples of how connected

ensembles of neurons are thought to implement brain processing.

1.1.1 Neurons as basic units for brain processing: facts and experimental techniques

Here, we describe how the neurons, which are the basic cellular units which compose the brain, can emit, transmit and receive signals. We then briefly expose the experimental techniques permitting to study neuronal activity, which we will refer to later in the dissertation. We first explain how neuronal activity can be recorded. We finally describe the techniques by which neurons may be artificially stimulated, and the limits of these techniques.

Basic mechanisms of single neuron function

Throughout this section, we will summarize basic facts about single-neuron dynamics. As a reference, we rely on [Gerstner et al. (2014)]. Neurons are cells possessing an excitable membrane. A sufficiently strong increase of the electric potential of this membrane, which can be induced by an injection of electrical charges inside the neurons, can trigger a positive feedback mechanism which actively amplifies the membrane potential increase. This leads to a prototypical excursion of the membrane potential followed by a reset of this potential to a baseline value. This prototypical time-course of the membrane potential is commonly referred to as a spike (or, equivalently, an action potential).

Each spike fired by a neuron is a signal which can trigger the release of a chemical, called a neurotransmitter, at specialized sites called synapses. Synapses are the connection points through which neurons can interact. More precisely, a neuron possesses a long tubular membrane extension from the cell body, which is called an axon. This axon then typically forms several branches, that terminate at different synaptic sites which are situated close to the membrane of receiving neurons. The receiving contact sites are usually situated rather close to the main body of the neuronal cell. These post-synaptic input sites may be regrouped on specialized neuronal extensions called dendrites [Llinas (2008)].

When a first (so-called “pre-synaptic”) neuron emits a spike, the depolarization of the membrane potential is transmitted along the axon. This causes the release of neurotransmitter molecules at the synaptic sites. These

neurotransmitter molecules can then diffuse to the membrane of the target (so-called "post-synaptic") neurons. Finally, the neurotransmitter molecules bind to post-synaptic membrane receptors. This triggers the transient entry of electric charges in the post-synaptic neuron. More specifically, the binding of the neurotransmitter molecules cause the direct or indirect opening of transmembrane proteins which then act as channels that specifically allow some types of ions to travel across the membrane.

Some neurons, called excitatory neurons, send excitatory neurotransmitters. These neurotransmitters trigger an increase in the membrane potential – referred to as a **depolarization** – of the post-synaptic neuron. The most prominent types of excitatory neurons are the pyramidal neurons, which are named after their shape [Spruston (2009)].

Other neurons are inhibitory: they send neurotransmitters which trigger a decrease in the membrane potential – referred to as a **hyperpolarization** – of the post-synaptic neuron. Most of the interneurons, which are small neurons primarily sending local connections, are inhibitory [Freund and Kali (2008)].

Depending on the nature of the receptor, the duration of the episode of charge entry after a pre-synaptic spike may vary. For instance, excitatory receptors such as those of the AMPA type (named after a molecule, the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid, that can bind to them) possess a fast time scale of one or two milliseconds. Other excitatory receptors, named NMDA receptors (for N-Methyl-D-aspartate, a molecule that can bind to them) have a longer time-scale of about a hundred milliseconds. Several time-scales also exist for inhibitory receptors. Finally, the electric charges coming from many synapses are summed in the post-synaptic neuron and they trigger changes of its membrane potential. This phenomenon generally involves a low-pass filtering due to the neuronal membrane properties, and a non-linearity. Finally, if the membrane potential of the post-synaptic neuron is sufficiently depolarized, a post-synaptic spike may be triggered in response to the input electrical charges received at the synapses.

Note that these points will be expounded more formally and quantitatively in the theoretical part of the dissertation.

We will now explain how neuronal activity may be studied through neuronal recordings.

Recording neuronal activity

The activity of neurons may be recorded through electrodes. Electrodes are devices that measure a difference of electrical potential, which relates to the difference in the density of electrical charges between two recording areas. In our case, one of these recording areas is a reference point (the ground) whose potential does not vary, and the other will be either the intracellular area of one neuron, or the extracellular area surrounding one neuron.

Intracellular recordings. A technique named "patch clamp" allows experimentalists to seal an electrode tip around a small hole in the membrane of a neuron [Moore (2007)]. Hence, the electrode can sense the intracellular potential of the neuron, which permits to record both the time-course of the potential below the threshold for spiking, and the spikes. This technique hence yields very precise data. However, one of its disadvantages is the requirement to form a stable seal with the neuron. Therefore, this technique is mostly employed in brain slices (i.e., "in vitro") and much less often in alive animals (i.e., "in vivo").

Extracellular recordings. It is also possible to insert electrodes in the extracellular medium surrounding the neurons. This confers the considerable advantage to permit recordings in awake, behaving animals. However, in this case, the recorded potential is only an indirect measure of the intracellular potential of the neuron. Hence, the signal-to-noise ratio is smaller, and it is only possible to reliably detect the changes of potential occurring during spikes. Further, given that different neurons are at different distances from the electrode, and given that different neurons can emit spikes of different shapes, different neurons are likely to yield signals of separable shapes and amplitudes. Hence, it is possible to classify the detected spikes in different clusters which putatively correspond to different neurons [Harris et al. (2000)]. This technique is referred to as spike sorting. Despite the obvious limitations of the approach, its reliability has been shown to be rather reasonable: between 70 and almost 100% depending on the the specific algorithm (or person...) used to classify spike shapes ([Harris et al. (2000)]). For instance, refinements of this technique involve the insertion of several electrodes, and the detection of a single neuron on several of these.

Until recently, technical limitations imposed to insert only a few such electrodes. Hence, typically, only a few neurons could be simultaneously recorded. Today, however, it is possible to insert a large number of fine electrodes and to record a hundred neurons simultaneously [Stevenson and Kording (2011)]. This is of importance to improve the understanding of neuronal computations, as they are thought to emerge from connected populations of neurons (as we will soon explain in more details).

We will now show how the characteristics and the function of the neuronal response to input currents can be investigated through artificial stimulations of the neurons.

Artificial stimulation of neurons

Several techniques can be used to stimulate neurons artificially. Experimentalist-controlled stimulations can indeed precisely inform about the dynamical response of single neuron to stimulation, and about the function of the neuronal activity in behaving animals.

Stimulation through intracellular current injections. When using the above-mentioned patch-clamp technique, it is possible to simultaneously inject charges into the neuron, and record the intracellular membrane potential. This permits to study in great details the input-output function of the neuron. In particular, this technique can be used to quantitatively fit a model for the dynamic neuronal response to rich non-stationary input currents [Mensi et al. (2012); Pozzorini et al. (2013)].

Extracellular stimulation. When using extracellular electrodes, it is also possible to inject electrical current. This will excite a small population of neurons situated close to the electrode tip. This type of stimulation is often used in awake, behaving animals. Indeed, the causal relation between an increased activity in the population of neurons situated close to the electrode tip and the animal's behavior can then be assessed (see [Hanks et al. (2006)] for an example). Note that the success of this approach relies on the fact that in some areas of the brain, neighboring neurons often share similar properties [Schall et al. (1995); Hanks et al. (2006)].

Optogenetic stimulation. Optogenetics is a new technique that was developed recently, and which permits to modulate the activity of populations of neurons in behaving animals [Fenno et al. (2011)]. This technique relies on making neurons artificially express some transmembrane ions channel proteins. This protein expression is controlled through genetic manipulation. The technique can be used with channels that are specific to either positively or negatively charged ions, and which can respectively induce an excitation or an inhibition of the targeted neurons. Importantly, the experimentalist can control the opening of a specific channel type by shining light at a specific wavelength. There are different techniques which permit to shine lights on the neurons of interest, which range from the insertion of optical fibers in the brain (to target deep brain structures) to the removal of the skull (to target upper cortical layers).

This technique has the considerable advantage to be able to target the stimulated neurons through both the restriction of the area receiving the light, and the expression of the above-mentioned channels. This expression can be controlled by injecting pieces of DNA composed of a part coding for the channel, and of a regulating element which conditions the expression of the DNA to the presence of a particular cell protein (called a transcription factor). Different neuron types (such as pyramidal neurons vs. interneurons, or neurons in some specific brain areas) express different types of transcription factors. Hence, the stimulation can be specific to such a genetically defined population of neurons. In addition, the DNA can also be engineered such that it is not expressed in the presence of a drug that can be fed or injected to the animal. As a consequence, it is possible to restrict the temporal window when channel expression can occur to a few hours. Finally, increased neuronal activity triggers the expression of a transcription factor (c-Fos), on which the expression of the light-activated channels can be conditioned [Liu et al. (2012)]. This can be used to specifically target a population of neurons which shows sustained increased activity during a certain behavior of the animal, or when the animal is placed in a given context.

Hence, optogenetic tools can be used to control increased or decreased activity to populations of neurons that either possess a specific transcription factor, or that are specifically and strongly activated in a given situation. There are however limitations [Ohayon et al. (2013)]. First, it may not be possible to target a desired population of neurons, because these neurons may neither differ genetically from the others, nor show sustained activity during a specific context that can be imposed on the animal to enforce channel expression through c-Fos. Second, in

general, the technique cannot be used to enforce a very precise intensity for the stimulation in all targeted neurons, as the intensity depends on both channel expression and light reception. Third, for large animals, there may be a difficulty to shine light on a sufficiently large number of neurons.

Optogenetics is nevertheless an important advance to study populations of neurons, which are thought to shape neuronal computations, as we will now briefly review.

1.1.2 Neuronal processing through connected populations of neurons

Different neurons may be connected in a feedforward fashion, hence forming a unidirectional chain of elements. An example of such a connectivity layout is the connection from the mammalian touch cutaneous receptors to the second-order touch neurons [Moayedi et al. (2015)].

Alternatively, the connections between neurons may be recurrent (i.e., with direct or indirect reciprocal connections), as for instance observed in the mammalian prefrontal cortex [Wang et al. (2006)].

We now exemplify how these connection schemes relate to different types of brain processing.

Sensory processing

The sensory areas of the nervous system, such as the primary visual cortex or the cuneate nucleus of the mammalian brain, receive and process information coming from biosensors, such as the retina or the skin touch receptors [Carandini (2012); Moayedi et al. (2015)].

The sensorial stages of neuronal processing are often a series of feedforwardly connected layers of neurons. We already mentioned the touch system [Moayedi et al. (2015)]. Another example, for which the feedforward property is approximately realized at a larger spatial scale, is the mammalian visual system. Indeed, the output neurons of the retina project to the thalamus, which in turn project the the primary visual cortex [Carandini (2012)].

The peripheral biological sensors often send complex spatiotemporal signals to the primary sensory areas (e.g. [Bialek et al. (1991); Johansson and Birznieks (2004)]). Hence, in this context, it is well accepted that the timing of the emitted spikes is crucial for successful signal transmission. This type of signaling is referred to as **temporal coding** [Panzeri et al. (2010)].

Cognitive processing: basic facts and classical modeling frameworks

Cognition involves the selection (or the selective combination and processing) of some relevant information among the diversity of external and internal signals received by the brain. This process allows animals to use external cues and internal representations to fulfill internal goals, such as survival [Koechlin et al. (2003); Donoso et al. (2014)]. Hence, the maintenance of a relevant item in working memory, or the monitoring of some dynamical properties of a stimulus that are relevant for an upcoming decision, are both cognitive processes. In the mammalian brain, the frontal cortical areas are generally thought to be the main drivers of cognitive computations [Koechlin et al. (2003)].

Experimental characterization of neuronal cognitive computations. Experimental recordings in awake, behaving animals have yielded hypotheses for the neuronal correlates of cognitive processes.

For instance, the accumulation of evidence during sensory-based decision-making have been linked to ramping, "integration-like" firing rate increases in some populations of neurons of the lateral intraparietal cortex [Huk and Shadlen (2005); Hanks et al. (2006); Churchland et al. (2011)].

In addition, the maintenance of memory items during a delay period have been correlated to a sustained, quasi-steady activity in some neurons of the frontal cortex [Funahashi et al. (1989, 1993); Procyk and Goldman-Rakic (2006)].

Successful theoretical modeling of neuronal cognitive computations through recurrent networks. Compared to other neuroscience fields, cognitive neurocience has been linked to modeling and theory rather early on (e.g., [Hopfield (1982)]). This may be explained by the fact that cognitive computations are complex processes whose macroscopic properties were naturally seen as emerging from the combination of the activity of a large

number of individual components. These components could not all be monitored simultaneously. Indeed, during decades, it has been impossible to record many individual neurons simultaneously, which limited the understanding of the trial-specific mechanisms leading to a behavioral output. Even though these recording limitations are now being overcome, the issue is only partially solved. Indeed, the challenge is now to make sense of the available complex, high-dimensional data sets. Hence, the need for a simplification through theory was rather obvious from the start and remains valid today.

In consequence, several popular models were proposed to account for the observed neuronal correlates of cognition. Interestingly, in these models, the recurrent properties critically shape the dynamics [Compte et al. (2000); Brunel and Wang (2001); Wang (2002); Machens et al. (2005); Wong and Wang (2006); Hopfield (2007); Cain and Shea-Brown (2012); Deco et al. (2013); Lim and Goldman (2013)]. Through these recurrent connections, these models are indeed able to reproduce critical features of the experimental cognitive-related neuronal responses. Hence, persistent activity [Compte et al. (2000); Brunel and Wang (2001)], as well as integration-like ramping activity [Wang (2002); Machens et al. (2005)], can both be explained by those models.

Networks for cognitive processing are classically thought of reading information through a spatial rate code, rather than a temporal code. In these simple models for cognitive processing, the final output of the network which will ultimately trigger behavioral changes is classically characterized by a (quasi)-stable state of activity. This final activity state is usually assumed to depend on the identity of the stimulated neurons, and/or on the number of input spikes received by the network. For instance, the identity of an item held in memory, or the identity of a chosen alternative, could be encoded through a high-activity state sustained by recurrent excitation in a population of neurons [Brunel and Wang (2001); Deco et al. (2013)]. Hence, this type of network is characterized by multistability. The state of elevated activity of one recurrent population can be triggered by a transient episode of increased firing in the excitatory inputs it receives. Hence, in this type of models, the putative impact of a temporal structure in the synaptic input is typically not investigated.

Furthermore, other popular models for memory and decision-making are the above-mentioned approximate integrator networks [Cain and Shea-Brown (2012); Lim and Goldman (2013)]. They can accumulate evidence, and hold items in

memory, by firing with a rate that is approximately proportional to the number of spikes received from their external input. Hence, by the intended design of these networks, they should have little sensitivity to the temporal structure of the received external synaptic input.

To summarize, for these cognitive models, the relevant signal is almost always assumed to be contained in the identity of the neurons which fire, and in the intensity of their firing. This instantiates a so-called **spatial rate coding** paradigm.

Therefore, the dynamics of the models that we described above has proven to be powerful to give insights about key aspects of cognitive processing, without the need to account for the role of temporal structure. In addition, the recurrence and the non-linearity of these types of network actually make it difficult to analyze how the temporal structure of the synaptic input could shape their dynamics [Gerstner et al. (2014)]. This helps explaining why the question of a possible function of the input's spike timings had been mostly overlooked in this context. Furthermore, the amplification of spike time noise during the steady-state activity of cortical networks has been used as an argument against the possibility of precisely timed patterns of spikes during cognitive computations [London et al. (2010)]. The authors concluded that a **temporal coding** paradigm, in which the temporal structure of the input is crucial for shaping the dynamics and the final state of the network, was therefore unlikely to underlie cognitive computations.

Finally, another factor which may have discouraged further investigations about this issue may be linked to the difficulty of defining temporal coding in a meaningful and non-trivial way [Panzeri et al. (2010)]. Indeed, even in the simple networks mentioned above, which can work without a crucial function of the input's temporal structure, the firing rates are dynamic. Therefore, a temporal modulation of the neuronal activity does occur in these models of cognitive processing. In this context, temporal structure can be seen as an epiphenomenon, and focusing on it could be considered as detrimental for reaching an understanding of the circuit's function. In addition, the real circuitry can obviously only be an approximation of the simple rate network models that were proposed for cognitive function. Therefore, some deviations from the simple framework sketched by the models could be seen as "bugs" rather than features, and focusing on them could again be considered as prejudicial for getting the big picture. For instance, concerning the neural integrator models, biologically plausible implementations [Wang (2002); Wong

and Wang (2006); Wong and Huk (2008); Lim and Goldman (2013)] possess a slow leak and a small non-linearity. Notably, the leak term, which implements a low-pass filter [Naud and Gerstner (2012b)], will lead to a modulation of the response of the network depending on slow temporal variations of its synaptic input. More precisely, this modulation occurs when the input's temporal variations are about as slow as or slower than the leak time-scale. However, this small leakiness is not assumed to play an important role in neuronal processing in the context of an approximate integrator network. Rather, the leakiness is a consequence of biological limitations. Hence, even though a purely theoretical, perfect integrator would be completely insensitive to its input's temporal structure, a sensitivity to the slow temporal variations of the input cannot be taken as an evidence against the integrator model. Rather, what matters is whether the major properties of the real network are consistent with an approximate integration.

In other words, a naive analysis which would merely report the presence of some temporal structure in the neuronal response is likely to not be very informative about the essence of the neuronal computation at stake.

1.1.3 The relevance of temporal structure for driving networks with cognitive function: an unanswered but nevertheless relevant question

In this context, why would one ask the question of the function of temporal structure during cognitive processing? The answer is simple: the above arguments do not exclude the possibility that a carefully designed study focusing on temporal structure could be insightful for understanding the computations at stake. First, while a network which implements an approximate integration should not – by definition – be sensitive to the fine temporal structure of its input, there is no reason to believe that the multistable networks are not sensitive to their inputs' spike times. On the contrary, the non-linearity of these multistable networks is actually likely to make them sensitive to their input's temporal structure, even though in general they are only fed simple step-like firing rate inputs. Interestingly, a recent study indeed showed that the input's temporal structure can robustly modulate the dynamics of such networks and could be used to control the probability that the network switches to a

different stable state [Dipoppa and Gutkin (2013b)].

Hence, evidence for a functional relevance of the input's fine temporal structure could be seen as arguing against the processing of this input by an integrator network. In addition, such evidence would be compatible with a functional relevance of the non-linear behavior of the **decoding** network (which processes the temporally structured input).

Despite the numerous possible pitfalls, we therefore feel that it is worth investigating whether or not the input's temporal structure could sizably shape the output of a network implementing cognitive computations. This could indeed be extremely insightful about the basic biological mechanism implementing the computation. Finally, this may in turn have a large impact on our understanding of the function played by this network for shaping the adaptation of the animal's behavior.

1.2 Objectives of the doctoral study

During the doctoral study, we first aimed at investigating to what extent the temporal characteristics of a neuronal signal fed to a network with cognitive function could be consistent with the hypothesis that this network behaves as an integrator (whose sensitivity to temporal structure is weak). To this end, we analyzed data from an area involved in cognition: the dorsal anterior cingulate cortex (dACC). This area is activated in a variety of contexts which require animals to adapt their behavior to dynamic environmental cues [Procyk et al. (2000); Procyk and Goldman-Rakic (2006); Quilodran et al. (2008); Hayden et al. (2011a,b); Sheth et al. (2012); Blanchard and Hayden (2014)]. A recent theory unified these findings by suggesting that dACC could transmit a signal which would specify an adapted behavioral strategy, and/or which would quantify to what extent it is worth allocating cognitive resources to update the behavioral strategy [Shenhav et al. (2013)]. Interestingly, the latter signal (referred to as "expected value of control") is a scalar, one-dimensional quantity which could naturally be encoded through different intensities of firing. This signal could in turn be easily decoded and maintained in memory by a downstream neural integrator network. In addition, the literature suggests that dACC activity is read out by the dorsolateral prefrontal cortex during cognitive processing [Procyk and Goldman-Rakic (2006); Rothé et al. (2011); Shenhav

et al. (2013). Interestingly, the dorsolateral prefrontal cortex is an area which has been shown to behave similarly to an integrator in some contexts ([Kim and Shadlen (1999), but see [Rigotti et al. (2013); Hanks et al. (2015)]).

Hence, it was relevant to consider and probe the possibility that dACC activity could be decoded by an approximate neural integrator, which would have a very weak sensitivity to dACC spike timing.

We therefore wished to test the presence of a temporal structure in dACC activity that would be functionally relevant. More precisely, we intended to assess this functional relevance in terms of improvement of the decoding of dACC activity during cognitive control, as well as in terms of correlation between dACC activity and future behavior of the animal.

A second important objective of the doctoral project was to propose a plausible neuronal network which could process dACC activity in a way that would be consistent with the conclusions of our data analysis. More precisely, the aim was to deepen the understanding of the mechanisms by which temporal structure could participate to shaping the dynamics of the network decoding dACC spike trains.

1.3 Road map of the dissertation

This introduction, which sketches the general approach taken during the doctoral work, constitutes **Part I** of the dissertation.

Part II reports data analysis results which show evidence in favor of a spike-timing sensitive, non-linear decoder of cognitive-control related discharges. This part of the dissertation corresponds to a rearrangement of a recently published article [Logiaco et al. (2015)]. In order to show the entirety of the results, we present this part of our research through a seamless and slightly enriched text, in which the relation of the results to modeling has been extended. This presentation of our data analysis incorporates the supplementary information of the published article. We note that we used the figures as they were made for this article. Many were supplementary figures, which were often made a posteriori to answer reviewer's comments. This implies that these figures often relate to different subsections of the new layout. We apologize for the inconvenience that this may cause during the

reading of the manuscript. In this part of the dissertation, we first introduce in more details the state-of-the-art knowledge about dACC signaling during cognitive control, as well as the definition of temporal coding (in chapter 2). We then describe our analysis methods in chapter 3. After this, we present our results in chapter 4. Finally, we discuss the implications for the function of the neuronal network(s) which process dACC signals in chapter 5.

Part III describes a simple analytical tool permitting to investigate the impact of the input's temporal structure on the dynamics and function of recurrent neuronal networks. This part starts with a preamble explaining our working hypothesis for the dynamics of the network processing dACC activity (in chapter 6). We also explain why the previously existing theoretical tools revealed insufficient to permit a satisfying analysis of such a network. This preamble is followed by an an introduction (chapter 7), which expounds the unfulfilled need for mathematical expressions describing the dynamics of networks of recurrently connected single-neuron models which can be fitted to neuronal recordings. These expressions have to account for the high variability of neuronal spiking during functional cortical activity, as well as for the strong adaptation properties of excitatory neurons. We then describe in chapter 8 the mathematical analysis we developed to fill this gap in the theoretical literature. After this, we present some tests for the accuracy of our analytical results, as well as some applications, in chapter 9. We mention how our new theoretical tool could be used to tackle in more details the question of the processing of dACC activity by a recurrent neuronal network. Finally, we discuss the novelty of our theoretical results in chapter 10.

Part IV concludes the dissertation. It summarizes how the doctoral work contributed to deepen the understanding of how the temporal structure of neuronal activity could be functionally relevant during cognitive computations implemented by recurrent neuronal networks.

Part II

Evidence for a spike-timing-sensitive and non-linear decoder of cognitive control signals

Chapter 2

Introduction: signals for behavioral strategy adaptation in the dorsal Anterior Cingulate Cortex

---

Cognitive control is the management of cognitive processes. It involves the selection, treatment and combination of relevant information by the brain, and it allows animals to extract the rules of their environment and to learn to respond to cues to increase their chances of survival [Koechlin et al. (2003); Ridderinkhof et al. (2004)]. Evidence strongly suggests that frontal areas of the brain, including the dorsal anterior cingulate cortex (dACC), are involved in driving this behavioral adaptation process. However, the underlying neuronal mechanisms are not well understood [Shenhav et al. (2013)].

2.1 Cognitive control is most often thought to be supported by long time-scales of neuronal processing

Most studies have focused on the number of spikes discharged by single dACC units after informative events occur. Other potentially informative features of the neural response, such as reproducibility of spike timing across trials, have typically been ignored. The reason for that may be the apparent unreliability of spike timing when observing frontal activity, which seems to be in agreement with theoretical analyses of the steady-state activity in recurrent networks [London et al. (2010)]. Also, cognitive processes often involve to hold

information in working memory, a process that can naturally be implemented through networks possessing a long time-scale (on the order of seconds, [Lim and Goldman (2013); Cain and Shea-Brown (2012); Wong and Wang (2006)]). Accordingly, most models of cognitive processing [Brunel and Wang (2001); Mongillo et al. (2008); Rolls et al. (2010); Cain and Shea-Brown (2012)] rely on stepwise firing rate inputs, therefore disregarding the potential impact of the finer temporal structure of the driving signals. In the specific case of dACC, a recent theory [Shenhav et al. (2013)] suggests that this area transmits a graded signal: the expected value of engaging cognitive resources to adapt the behavior. This signal has to be remembered from the moment when the current behavioral policy appears to be improper until the moment when a more appropriate strategy can be implemented. Hence, a simple neural integrator [Churchland et al. (2011); Cain and Shea-Brown (2012); Lim and Goldman (2013); Bekolay et al. (2014)], which by construction is insensitive to spike timing, would be well suited to decode and memorize this signal. This neural integrator could be implemented by the lateral prefrontal cortex ([Kim and Shadlen (1999), but see [Rigotti et al. (2013); Hanks et al. (2015)]), which is a plausible dACC target during behavioral adaptation [Procyk and Goldman-Rakic (2006); Rothé et al. (2011); Shenhav et al. (2013)].

2.2 A gap in the literature concerning the processing time-scale during cognitive control

Some other brain regions that are not primarily involved in cognitive control are however known to be sensitive to both the timing [Bialek et al. (1991)] and the spatial distribution [Aronov et al. (2003)] of spikes within their inputs. These features may improve information transfer between neurons through, for instance, coincidence detection [Rudolph and Destexhe (2003)].

It is worth noting that, in frontal areas (including dACC) involved in behavioral adaptation, several studies reported the presence of a temporal structure in neuronal activity [Shmiel et al. (2005); Sakamoto et al. (2008); Benchenane et al. (2010); van Wingerden et al. (2010); Buschman et al. (2012); Narayanan et al. (2013); Totah et al. (2013); Stokes et al. (2013); Womelsdorf

et al. (2014)]. This opens the question of whether fine spike temporal patterns could be relevant for cognitive control. However, the current observations are not sufficient to conclude about the relevance of this temporal structure for downstream stages of neuronal processing, and for the decision about future behavior. Indeed, to the best of our knowledge, there exists no study comparing the reliability and correlation with behavior of spike count and spike timing in individual frontal neurons during a cognitive task. Comparing spike count vs. spike timing sensitive decoders is central to the general view of temporal coding [Panzeri et al. (2010)]. In this framework, temporal coding can be defined as the improvement of information transmission based on sensitivity to spike timing within an encoding time window [Panzeri et al. (2010)]. In the case of discharges related to behavioral adaptation, which do not in general transmit information about the dynamics of an external stimulus, this encoding time window can be taken as the time-interval of response of the relevant population of neurons [Panzeri et al. (2010)].

In fact, some temporal structure can be present within this encoding window while still not improving decoding, because spike timing and spike count can carry redundant information [Oram et al. (2001); Chicharro et al. (2011)]. In addition, realistic neuronal decoders are likely to be unable to be optimally sensitive to all statistics of their inputs. In particular, neurons and networks are likely to trade off temporal integration with sensitivity to spike timing [Rudolph and Destexhe (2003)]. This also participates to explaining why, even in the presence of temporal structure, the decoding strategy leading to highest information (among those that can plausibly be implemented during neuronal processing) may be temporal integration [Chicharro et al. (2011)].

Further, the temporal structure can be informative but still fail to correlate with behavior, suggesting that downstream processes disregard it and, instead, rely solely on neural integration (as reported in [Luna et al. (2005); Carney et al. (2014)]). This may reflect that the constraints on decoding strategy of downstream areas are mainly not on the maximization of the discriminability of the studied responses (usually, single-unit response to a limited stimulus set). Information might not be a limiting factor as downstream areas have access to many presynaptic neurons with either quite uncorrelated noise that can cancel when their responses are pooled, or with correlations that do not impair information transmission [Moreno-Bote et al. (2014)]. Instead, the constraints

could be the generalization of the computation to different types of stimuli, or the difficulty of learning the decoding network's connectivity.

Hence, it is not trivial to determine whether the temporal patterning of spike trains in frontal areas is actually relevant for the neuronal processes mediating behavioral adaptation.

2.3 Investigation of the nature of the decoder for behavioral adaptation signals in dorsal Anterior Cingulate Cortex

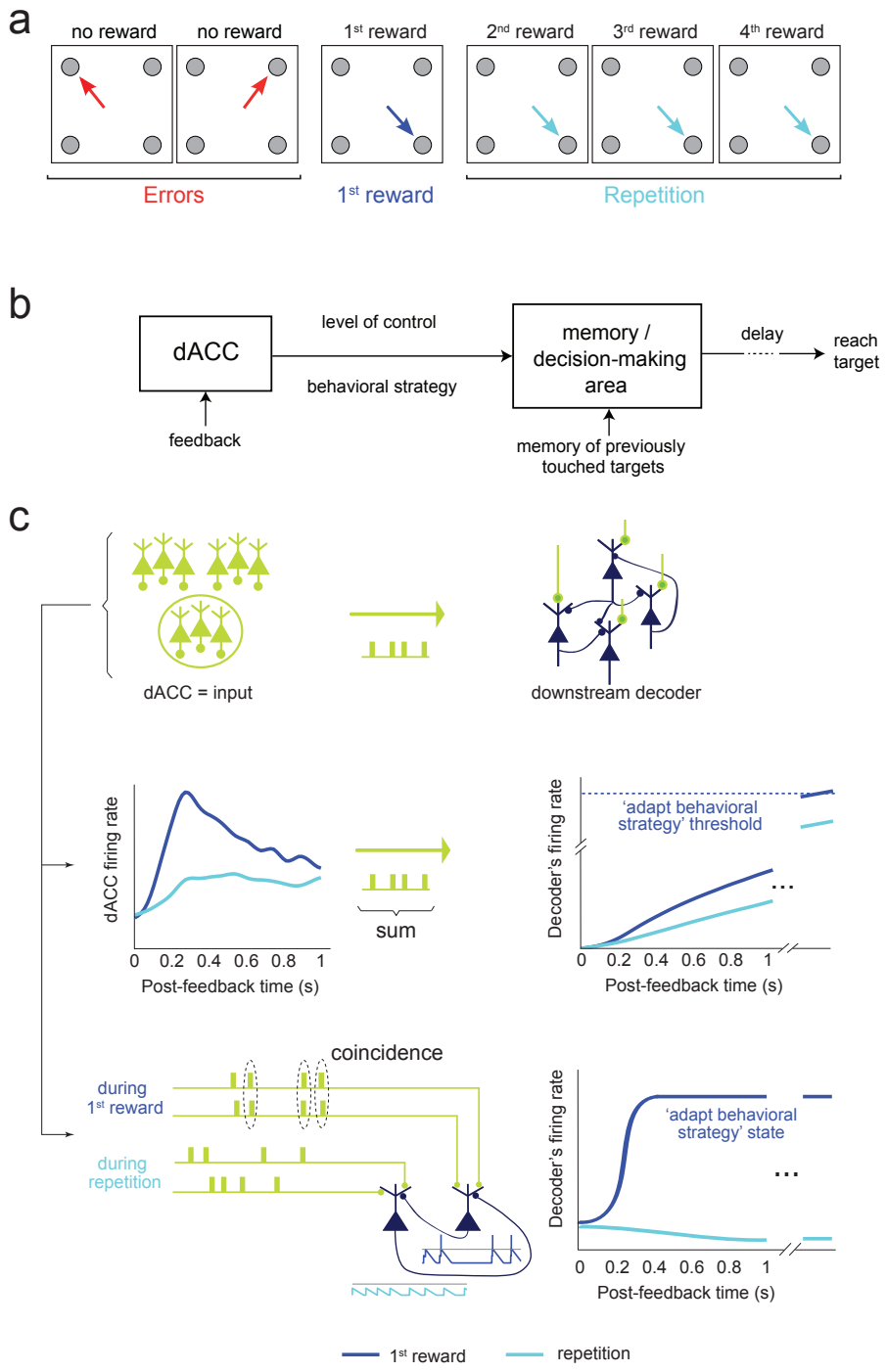


Figure 2.1: *Task and proposed neural mechanisms.* **(a)** During exploration, monkeys had to find, by trial-and-error, which of 4 targets resulted in a reward. After receiving the 1^{st} reward, monkeys entered a repetition period and received additional rewards by touching the same target. **(b)** Plausible dACC role in the task [Quilodran et al. (2008); Shenhav et al. (2013); Khamassi et al. (2013); Ullsperger et al. (2014)]: it processes feedback information (error or reward) to signal a behavioral strategy (either exploration, or switch toward repetition, or repetitive behavior). It would also signal the adaptive value of updating the behavioral strategy ("level of control"). A downstream area would combine dACC signals with a memory of previous choices to decide which target to choose next. **(c)** Spike count vs. timing sensitive decoding of dACC signals. Middle: a neural integrator decoder [Kim and Shadlen (1999); Cain and Shea-Brown (2012); Lim and Goldman (2013)] responding with a firing rate proportional to the sum of input dACC spikes. The decoder maintains a memory of past inputs and can store a continuum of level of control values. dACC neurons firing preferentially during either errors, or 1^{st} rewards, or both [Quilodran et al. (2008)] could project to different neural integrators. Bottom: an example of spatiotemporal decoder that is sensitive to the temporal structure of dACC spike trains and implements a memory. The connections between neurons create two stable states, with high and low firing [Brunel and Wang (2001); Dipoppa and Gutkin (2013b)]. The high activity state sustained through recurrent connections signals the need to adapt behavior. This decoder would be sensitive to its input's temporal structure, with some patterns favoring the transition to, and/or stability of, the high activity state [Dipoppa and Gutkin (2013b)]. This simplified scheme illustrates how temporal coincidences in the input may favor the discharge of downstream neurons.

Here, we address the issue of temporal coding of behavioral adaptation signals emitted by dACC neurons. We use recordings from monkeys engaged in a trial-and-error learning task [Quilodran et al. (2008)], in which performance relied on reward-based decision making and behavioral adaptation (Figure 2.1 (a)).

The task consisted in finding by trial and error which one of 4 targets was rewarded. Each trial led to the touch of a target and a feedback: a reward if the touch was correct, nothing otherwise. In each block of trials (i.e. a problem), monkeys first explored the different targets in successive trials. The first reward indicated discovery of the correct response. Then, a period occurred when the monkeys could repeatedly touch the correct target in 3 successive trials to exploit and receive additional rewards. The firing rate of single dACC units was previously shown to increase at feedback time during either exploration, or repetition, or when switching between those two states [Quilodran et al. (2008)]. Hence, dACC neurons may signal whether and/or how behavior should be adapted. In this context, we probe the putative structure and function of a downstream neuronal network decoding dACC feedback-driven signals. To do so, we investigate to what extent the temporal structure of dACC spike trains, during post-feedback firing, could improve information transmission and predict behavior (Figure 2.1 (b)). Assuming a neural integrator decoding scheme, the downstream network would compute and maintain the memory of the need for behavioral adaptation on the basis of the number of spikes emitted by dACC (Figure 2.1 (c), middle). This decoding network is therefore insensitive to its input's temporal structure at a finer time scale than the approximate integration (i.e., memory) time scale.

Alternatively, the downstream network could be sensitive to the spatiotemporal structure of dACC activity (Figure 2.1 (c), bottom). For instance, temporal coincidences in the afferent dACC signals could favor the switch to, and maintenance of, a high-activity state in the downstream network to encode behavioral adaptation [Dipoppa and Gutkin (2013b), see also Gutkin et al. (2001); Dipoppa and Gutkin (2013a)]. Note that other mechanisms could also explain the sensitivity of the downstream memory/decision network to temporal structure (for another example, see [Szatmáry and Izhikevich (2010)]). Notably, any network for which some non-linearity in the neuronal combination of synaptic inputs sizably shapes the output signal would be expected to have some sensitivity to the timing of input spikes. In the theoretical part of this dissertation (**Part III**), we will tackle in more details the question of the nature

of a decoding network that could hold items in memory, be sensitive to its input's temporal structure and be consistent with the experimental literature. Here, we focus on determining to what extent the characteristics of dACC feedback-related discharges could be consistent with a decoding by a network behaving as an approximate integrator.

We will actually show evidence suggesting that this is not the case. Instead, our analyses appear consistent with a non-linear spatiotemporal decoding of dACC activity.

First, we show that there are informative temporal patterns in single units that can support a larger reliability of a plausible spike-timing sensitive decoder, compared to a neural integrator. We found an optimal decoding time scale in the range of 70-200 ms, which is much shorter than the memory time-scale required by the task. The larger reliability of spike-timing sensitive decoding appeared to be supported by the combination of a large spike count variability, and a presence of informative temporal correlations between spike times.

Second, we show that some spike coincidences across jointly recorded neurons are advantageous for decoding. However, the informative spike times appear heterogeneous and distributed over the neuronal population, suggesting that downstream neurons could benefit from a non-linear spatiotemporal integration of inputs.

Finally, we describe a new method to evaluate to what extent dACC temporal patterns can predict the behavior of monkeys comparatively to spike count. Importantly, using this new method, we find that deviations from a prototypical temporal pattern sizably predict an increased response time of the monkeys.

Chapter 3

Methods for analyzing dorsal Anterior Cingulate Cortex activity and monkeys' behavior

---

In this chapter, we describe:

- the methods concerning the collection of the data (which was made in E. Procyk's laboratory, by R. Quilodran and M. Rothé), as well as the selection of the analyzed data, in section 3.1
- the methods (taken from the literature) used to investigate the coding properties of dACC spike trains, in section 3.2
- the methodology we used to verify that spike-sorting artifacts were unlikely to affect our results, in section 3.3
- a simple method we developed to analyze the monkey's eye movements from the X-Y position of one eye, in section 3.4
- a methodology we developed to investigate the relation between temporal patterns of spikes and the monkey's behavior, in section 3.5
- the classical statistical tests we used during the analysis, in section 3.6

3.1 Experimental methods

3.1.1 Electrophysiological recordings

Two male rhesus monkeys were implanted with a head-restraining device, and neuronal activity was recorded by 1 to 4 epoxy-coated tungsten electrodes (horizontal separation: 150 μm) placed in guide tubes and independently

advanced in the dorsal bank of the rostral region of the cingulate sulcus. Recording sites were confirmed through anatomical MRI and histology [Quilodran et al. (2008); Rothé et al. (2011)]. Extracellular activity was sampled at 13 kHz and unitary discharges were identified using online spike sorting based on template matching (MSD, AlphaOmega). All experimental procedures were in agreement with European, national, and local directives on animal research.

3.1.2 Problem solving task and trial selection

Monkeys had to find, by trial-and-error, the rewarded target among 4 targets presented on a touch screen (Figure 2.1 (a)). To begin a trial, the animal had to touch a central item (lever), which triggered the appearance of a fixation point. After 2 s of gaze fixation, the 4 targets appeared simultaneously. At fixation point offset, the animal had to select a target by making a saccade toward it, fixate it for 0.5 s, and touch it following a GO signal (i.e. all targets bright). All targets dimmed at the touch, and switched off after 0.6 s. Reward (fruit juice) was delivered if the correct target was selected, otherwise no reward occurred. Throughout this dissertation, we define a trial as the period of time between the touch of the lever and 1 s after the reception of a feedback (either error, or 1^{st} reward, or repetition reward). In addition, we call task epoch the time interval between 1 ms and 1 s after the reception of a given feedback. After a feedback, a time break of 2 s was imposed before starting a new trial. Any break in gaze fixation or touch within a trial led to resuming the sequence at the lever touch. Note that we did not consider that this started a new trial. In case of an incorrect choice, the animal could select another target in the following trial, and so on until the discovery of the rewarded target (hence, ending an exploration period). The correct target remained the same in the following trials, allowing the animal to exploit the rewarded action (during a repetition period). We define a problem as the block of trials associated with one rewarded target location. A flashing signal indicated the end of repetition and the beginning of a new problem (the new rewarded target had a 90% probability to be different from the target rewarded in the preceding problem). In a given problem, the reward size was constant; and within a session, up to two different reward sizes could be given. In 90% of problems the repetition period lasted 3 trials after the 1^{st} reward, whereas in 10% of problems 7-11 repetitions could occur. Repetition trials beyond the 3rd one

were excluded from the analysis to avoid possible surprise effects. At the time of recordings, the task was well known: monkeys only failed to repeat the correct touch in one of the trials following the discovery of the rewarded target in around 1% of problems. Then, both the incorrect touch and the following trials were discarded from analysis, but previous trials were kept. As previously reported [41], monkeys might be able to infer the rewarded target after 3 non-redundant errors, i.e. the 3rd error would systematically trigger a switch to repetition. Therefore, only 1^{st} and 2nd erroneous touches as well as 1^{st} rewards preceded by less than 3 errors were included in the analysis. For the repetition period, we selected all correct trials that followed a search with up to 3 preceding search errors.

3.1.3 Analyzed units

For monkey P, all recorded units were used. For monkey M, only units showing a significant response to at least one event (either error, or 1^{st} reward, or repetition reward, or fixation breaks) were used (TEST 1 in [Quilodran et al. (2008)]). The mean and standard deviations of the baseline firing rate (taken from -600 to -200 ms before feedback onset) were computed. Units with a change of firing rate of magnitude higher than 5 standard deviations of the baseline within more than six 10 ms bins between +60 and +800 ms of at least one event were selected. Note that this test cannot favor temporal coding in any way. This selection allowed us to focus on a reasonable number of neurons to analyze (in terms of computing time and statistical power, as there was a much larger number of recorded units in monkey M).

3.2 Methods for investigating the coding properties of spike trains

We will first present the methods for decoding dACC spike trains which inform about which (plausible) decoding network would be better suited to extract information from dACC spike trains. These methods can be found in subsection 3.2.1.

Then, we will present the methods for characterizing more in details the

statistics of dACC temporal structure that are useful for the (optimized) decoder. This method permits to distinguish between the contribution of a time-dependent firing rate, and the contribution of temporal correlations. These statistics can be linked to the biological mechanisms within dACC which are shaping the neuronal signal. These methods can be found in subsection 3.2.2.

3.2.1 Decoding dACC activity with a spike train metrics

Neuronal circuits can either detect coincident depolarizations due to spatiotemporally structured inputs, or loosely integrate all incoming inputs during a given task epoch (see Figure 2.1 (a,b), [Rudolph and Destexhe (2003); Cain and Shea-Brown (2012)]). Our analysis of dACC activity sought an unambiguous post-synaptic signaling of the task epoch during which dACC spike trains were emitted. This decoding approach is functionally relevant because different task epochs must result in different adaptations of the behavioral strategy in order to optimize performance.

When functioning in a coincidence detection mode, a post-synaptic neural decoder might discharge specifically to a given task epoch if its input spike trains would have a spatiotemporal structure more different between task epochs than within this epoch. Alternatively, a downstream neural integrator might become selective for task epochs by receiving inputs from neurons that fire more in one task epoch (see Figure 2.1, Figure 3.1).

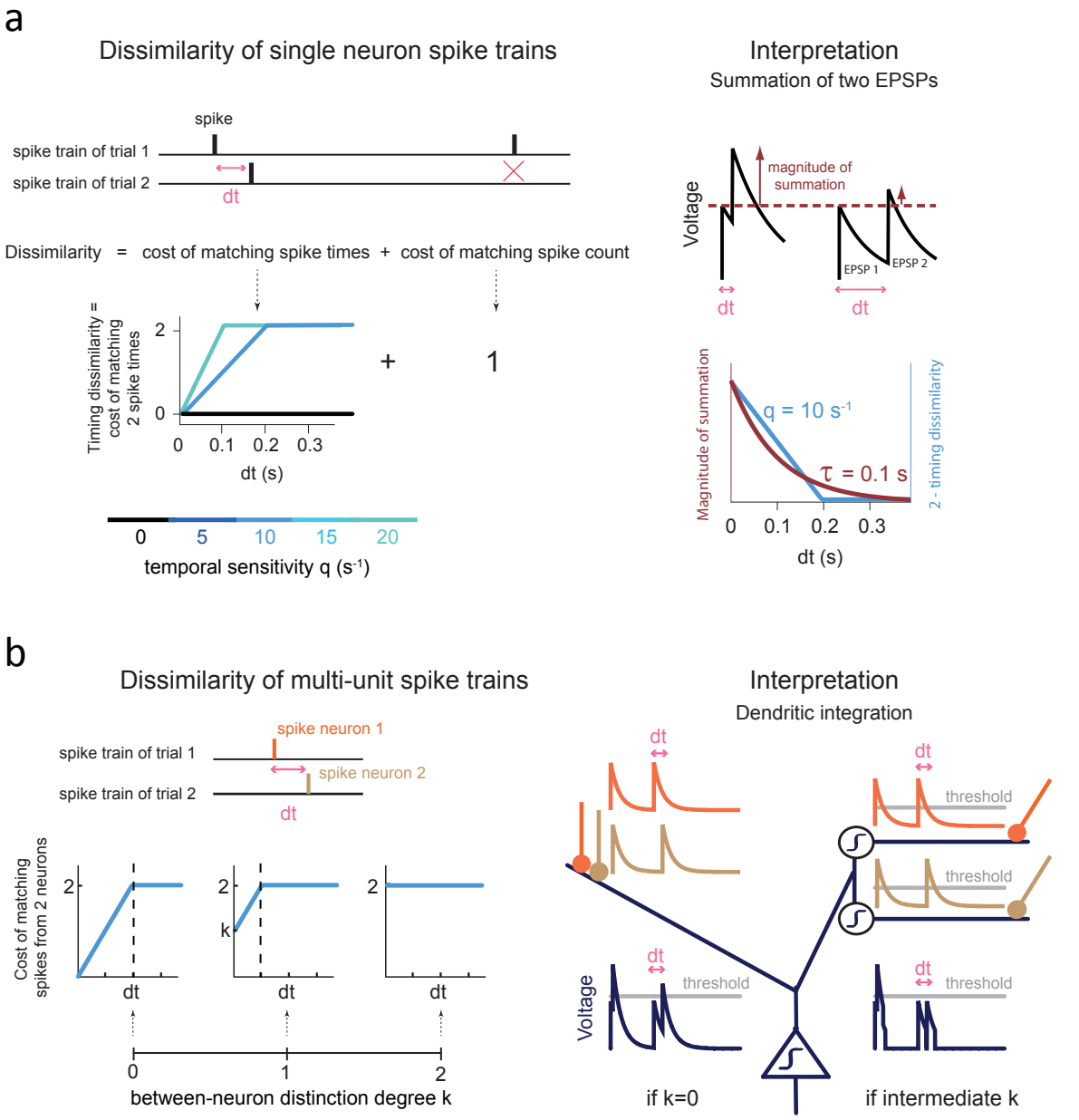


Figure 3.1: *Decoding method.* **(a)** Dissimilarity of single neuron spike trains. *Left*: the dissimilarity is the sum of the costs of matching spike times and cancelling the spike count difference [Victor and Purpura (1996)]. The cost of matching spike times depends on the parameter q (temporal sensitivity). When $q = 0\ s^{-1}$ (black curve), the dissimilarity only reflects the difference in spike count. For $q > 0\ s^{-1}$, the dissimilarity increases as q times the interspike interval dt before saturating at 2. *Right*: Each value of $q > 0$ can be related to a given time scale of Excitatory Post Synaptic Potentials (EPSPs, here taken as simple exponential traces: up). Indeed, decoding with this q value and decoding by summation of these EPSPs both lead to a similar sensitivity to spike timing. For instance, $q = 10\ s^{-1}$ corresponds to a 0-200 ms range of dt for which the dissimilarities are smaller than 2 (the maximum). This can be matched to the range of dt with efficient summation of 2 EPSPs decaying with 100 ms time scale (see section 3.2.1). The 0-200 ms range of dt therefore gives rise to temporal coincidences. **(b)** Dissimilarity of multi-unit spike trains. *Left*: computation of the dissimilarity between two spike trains, each of which contains spikes from 2 neurons [Aronov et al. (2003)]. The dissimilarity depends on the parameter k, which determines the degree of distinction between the 2 neurons. The cost of matching 2 spikes is increased by an amount k if the 2 spikes were emitted by 2 different neurons. As k increases the matching of spikes emitted by the same neuron is favored. For higher values of k, there is a smaller range of between-neuron interspike intervals leading to dissimilarities smaller than 2 (i.e. leading to a temporal coincidence). *Right*: higher values of k can be related to larger non-linearities in dendrites (here taken as thresholds and symbolized by a step within a circle). In the left dendrite, there are no non-linearities: the synapses are close and the depolarizations due to synaptic inputs can be directly summed and trigger firing (by crossing the threshold of the soma twice). This can mirror a maximal between-neuron summation, i.e. k=0. Conversely, in the right dendrite, the two synapses are on different sub-branches which both possess a threshold non-linearity. These thresholds (below which the synaptic currents are not transmitted to the soma) can prevent effective summation for large interspike intervals (second spike pair). This can mirror decoding with intermediate k values, causing only smaller interspike intervals to be associated with small dissimilarities between neurons (i.e. temporal coincidences).

The efficiency of a decoding strategy can be assessed by quantifying how dissimilar spike trains are, within and between categories, in terms of either (spatio)temporal structure or spike count. Within the theoretical framework named spike train metrics, the distance or dissimilarity between two spike trains is measured as a function of both the importance of spike timing [Victor and Purpura (1996)] and the spatial distinction between the activity from different input neurons [Aronov et al. (2003)].

Single-unit spike train metrics

The distance $d(s, s')$ between two spike trains s, s' is defined as the minimal cost to transform s into s' [Victor and Purpura (1996)]. This transformation consists in using one of the three following steps sequentially:

- adding a spike, for a cost of 1;
- deleting a spike, for a cost of 1;
- changing the time of a spike by an amount dt, for a cost $q \cdot dt$, where q is a free parameter that determines the importance of spike timing (also named timing sensitivity throughout the paper).

When $q = 0\ s^{-1}$, there is no cost for changing the timing. Consequently, the distance $d(s, s')$ corresponds to the absolute spike count difference between the two spike trains. As q increases, changing the timing of spikes becomes more and more costly. Thus, a small distance $d(s, s')$ implies that s and s' have spikes that match in time, i.e. the temporal structure must be conserved. Two spikes from s and s' may be moved to be matched if they are separated by at most $2/q$ second. Otherwise, it is less costly to delete the first spike and reintroduce a new matching spike, for a total cost of 2. Therefore, $2/q$ gives the maximal between-trial interspike interval for which timing is accounted for.

Multi-unit spike train metrics

A multi-unit spike train is defined as the pattern of discharges from different neurons observed in a given trial, each spike being labeled by the identity of the neuron that emitted it. To compute the distance $d(s, s')$ between two multi-unit spike trains s, s', two parameters can be considered: the timing sensitivity q, and the degree of distinction k between spikes from different neurons [Aronov et al.

(2003)]. For example, if two neurons emit spike trains with statistically identical temporal structures and fire with uncorrelated noise, then pooling their responses can be better for decoding. Conversely, if two neurons emit opposed signals (for instance an increase vs. a decrease of spiking in a given task epoch), then it is important to distinguish between them to maximize information. The distance $d(s, s')$ between two multi-unit spike trains is defined as the minimum cost to transform s into s', by using the steps previously described, with the additional possibility to change the identity of the neuron that fired a given spike, for a cost k. If $k = 0$, the identity of neurons does not matter at all. If $k \geq k_{max} = 2$, the responses are never matched between neurons, because removing a spike from a given neuron and replacing it by a spike from another neuron at the correct time is less costly. In general, two spikes from two different neurons may be matched if they are separated by less than $\frac{(2-k)}{q}$ second —so only very coincident spikes are matched for intermediate k values.

Classification

A leave-one-out process was used to classify a given spike train s into the task epoch E producing the most similar responses to s. The distance between s and the activity produced during E was defined as the median of the pairwise distances between s and any (other) spike train $s' \in E$. Therefore, one spike train s was predicted to belong to the task epoch E that minimized $median\left(d_{q,k}(s, s')\right)_{s' \in E,\, s' \neq s}$.

Note that we also ran a decoding analysis of dACC activity by using a small-distance biased classification algorithm orginally proposed by [Victor and Purpura (1996)] ($z = -2$ in their eq. 5, i.e. the distance between s and the activity produced during E is $\left(\langle (d_q(s, s'))^{-2} \rangle_{s' \in E,\, s' \neq s}\right)^{\frac{1}{-2}}$). We did not retain this method because (i) it hinders classification based on spike count decoding, and (ii) it leads to an overall decrease of the number of significant units and of the information (all analyzed single units, signed-rank test on $max_q(< I >_t)$, all $p_s < 10^{-5}$). These effects are likely to be related to the frequent occurrence of zero pairwise distances in our dACC data set (due, for instance, to two empty spike trains or, for $q = 0$ s^{-1}, to two spike trains with the same spike count). Although the occurrence of zero pairwise distances was more frequent within task epochs, given the high variability of our data (which we will show in Figure 4.10), it was also possible between task epochs. With the

small-distance biased classification, the presence of at least one zero pairwise distance in both epochs triggered a chance-based clustering of spike trains, irrespective of the 0-distance frequency in the two task epochs. Despite the lower classification power of this method, it leads to identical modulation of the classification performance by (q, k) as the median-based classification (results for the single-units classification will be shown in Figure 4.3). In general, for our very variable data, it is likely that any classification relying on outliers would be less efficient than a classification relying on a robust central value such as the median.

A confusion matrix was built, in which the entry N_{ij} on line i and column j was the number of spike trains coming from task epoch i and predicted to belong to task epoch j. If a trial was equally distant to several epochs, the fraction $\frac{1}{N_{closest\ epochs}}$ was added to all these epochs. The information I_{raw} in the confusion matrix was:

$$I_{raw} = \frac{1}{N} \sum_{i,j} N_{ij} \cdot ln\Big(\frac{N_{ij} \cdot N}{\sum_k N_{ik} \cdot \sum_l N_{lj}}\Big) \tag{3.1}$$

with $N = \sum_{i,j} N_{ij}$. This corresponds to the mutual information between the actual classification of trials and the classification that one would get if the prediction were perfect. Hence, I_{raw} is always maximal for perfect prediction, though the absolute maximum value depends on the balance of number of trials between the two task epochs. We finally computed a normalized information I_{norm} by dividing I_{raw} by its maximal (perfect prediction) value:

$$I_{norm} = \frac{I_{raw}}{-\frac{1}{N} \sum_i \Big(\Big(\sum_j N_{ij}\Big) \cdot ln\Big(\frac{\sum_j N_{ij}}{N}\Big)\Big)} \tag{3.2}$$

Note that this measure has the advantage of intrinsically accounting for the distribution of the number of data points in different categories to be classified, which is not the case of some other measures of classification performance, such as percentage of correct [Sindhwani et al. (2004)]. This was important in our case because there were much less 1^{st} reward or errors trials compared to repetition trials (see Table 3.1).

To test whether classification was above chance, trials were randomly permuted between task epochs, and two groups were recreated (with the same number of trials as the original task epoch clusters). The information content associated to the shuffled groups was then computed. The process was repeated 1000 times, leading for each q or [q,k], to 1000 values of information under the null hypothesis

		Behavioral adaptation	Repetition
Monkey M	1st reward discrimination	30 (16-40)	97 (62-130.25)
	Error discrimination	38 (27-47)	88.5 (68-120)
Monkey P	1st reward discrimination	17 (14-21)	60.5 (50-69)
	Error discrimination	27 (21-32)	59 (47.5-71)

Table 3.1: Median (and 25^{th} and 75^{th} percentile) number of trials for single-units that were selected as significant. For the paired analysis, trial numbers were similar, with exceptions when the two waveforms were jointly reliable only during a subpart of the recording (leading to slightly less trials).

that the discrimination between groups is due to random similarities between any two spike trains. The information analysis was done on increasing time windows, starting 1 ms after the onset of the feedback (to avoid pump-driven artifacts). The first window lasted until 50 ms post-feedback, and was incrementally increased to 600 ms by 50 ms steps, and then up to 1 s by 100 ms steps. The higher resolution for smaller windows allowed the time course of fast initial transient to be evaluated. We computed the maximum (over q or [q,k]) number N_w of consecutive windows for which the information was strictly larger than the 95^{th} percentile of the 1000 sets of permuted data. The same process was repeated for each set of permuted data, relative to the remaining 999 permuted sets. A neuron (or a pair) was considered as significant if N_w was strictly larger in the actual data than in 95% of permuted data. This process did not favor a given value of q or k, and could select neurons/pairs of neurons with different information time course. Also, it allowed us to exclude neurons with very unreliable activity, which would act as "noise" during the subsequent analyses.

The information estimate is, in general, biased when only finite data is available. However, because the spike train metrics method makes the assumption that spike trains within one task-epoch appear more similar to one another than spike trains taken from two different task-epochs, it is globally less likely to generate the huge finite sample positive bias observed with the "raw" binning method [Victor (2005)]. Because classical analytical formulae for bias estimation cannot be applied to the case of the confusion matrix [Victor and Purpura (1996)], the bias was estimated empirically as the mean information computed in 1000 data sets created by randomly permuting the trials between task-epochs (as in [Saal et al. (2009)]). This bias estimate, which was usually very small, was subtracted from the information estimate in the original data.

In rare cases when slightly negative values were reached after bias-sustraction, the final information value was set to 0. Note that we verified that the q_{opt} found for the 1^{st} reward vs. repetition classification was identical with or without bias correction, even though this classification had the smallest number of trials and could therefore be more sensitive to finite-sample effects. More generally, we assessed the possible remaining presence of a bias by computing for each neuron (or pair of neuron) the minimum trial number over task-epochs $N_{trial\ min}$. We then compared different statistics related to information (e.g. increase in information thanks to temporal sensitivity, gain in information during paired decoding, ...) between neurons (resp. pairs) with $N_{trial\ min}$ that was higher vs. lower than the median. While several factors may cause a difference between the group of high and low trial number (such as behavioral differences between sessions of different durations, ...), a finite-sample bias would be expected to have a very specific impact on the statisitical measurements. Indeed, a given effect may result from a bias if, consistenly in the two monkeys, the effect would decrease in the high trial number group and if this effect would be smallest in monkey M (which had the highest trial number, see Table 3.1). This pattern was never observed, arguing that our results are very unlikely to reflect a finite-sample bias.

The different parameter values were compared after bias correction. For each q or [q,k] and for each significant neuron, the temporal evolution of information values was summarized by taking the mean information over 10 analysis windows of increasing durations (ending from 100 ms to 1 s post-feedback onset, by steps of 100 ms, favoring neither early nor late information). We refer to this quantity as time-averaged information ($\langle I \rangle_t$, see Table 3.3 for a definition) in the dissertation. Computing the time-averaged information is equivalent to averaging over delays before a decision is made by the animal. Finally, a non-parametric Friedman ANOVA was used to compare the time-averaged bias-corrected normalized information as a function of different q or [q,k], with Tukey's honestly significant difference criterion correction for multiple comparisons. Note that there can be slight differences in the rankings of (q,k)-values between the mean-information time-course and the Friedman anova test. Indeed, the mean is more sensitive to outliers with large values, while the average rank used during the Friedman test is determined by the consistency (over neurons) of the within-neuron rankings of $\langle I \rangle_t$ between different (q,k)-values.

Interpretation of the classification as a downstream decoding network and non-triviality of the timing-related information improvement

The classification algorithm described in the previous section can be related to the performance of different dowstream neuronal circuits (see Figure 3.1). Indeed, the channels and membrane properties of single neurons can be approximately described by decaying filters (on the order of ms to hundreds of ms) of input spike trains [Gerstner and Kistler (2002)]. In addition, the neuronal network's architecture can create decays on much longer timescales, or even quasi-perfect integration, which may implement short-term memory [Seung et al. (2000); Lim and Goldman (2013)].

When the downstream neuronal network acts as an integrator, it effectively 'sees' input spike trains through their spike-count, and would perform a classification tantamount to the metrics with $q = 0\ s^{-1}$.

For $q > 0 s^{-1}$, the metrics is better interpreted through the equivalent similarity between spike trains. For any pair of spikes separated by an interval $\delta \geq 0$ and associated with a Victor and Purpura cost (or dissimilarity) $d(\delta)$, we can define the similarity $S = D_{max} - d(\delta)$. $D_{max} = 2$ is both the maximum dissimilarity between two spikes and the sum of the costs of removing a spike and of reinserting a new spike at the right time (see Figure 3.1 (a)). Hence, for $\delta \leq \frac{2}{q}$, $S(\delta) = 2 - q\ \delta$, and else $S = 0$. This similarity can be related to the maximal depolarization reached through the summation of two excitatory post-synaptic potential (EPSPs) that would be caused by the two compared spikes. Indeed, if we take the (plausible) choice of an exponential synaptic trace $A \exp(-\frac{t}{\tau})$ (for a post-spike delay $t > 0$), we can notice that the maximal depolarization reached after summation of the two filtered synaptic traces is $A + A \exp(-\frac{\delta}{\tau})$. We can finally define an 'excess depolarization' E above a baseline (here, the depolarization reached with a single spike): $E(\delta) = A \exp(-\frac{\delta}{\tau})$. The functions $S(\delta)$ and $E(\delta)$ have similar shapes and may be matched; in particular, we can equate:

- the maximal amplitudes of S and E: $A = D_{max} = 2$
- the integrals of S and E: $\int f(\delta) d\delta = A\tau = \int S(\delta) d\delta = \frac{2}{q}$

In other words, the (synaptic) decaying time-scale τ can be matched to $\frac{1}{q}$ [Victor and Purpura (1997); see also van Rossum (2001); Paiva et al. (2010) for related ideas]. For paired spikes, the more similar the two spikes are

according to the Victor and Pupura distance, the more excited would a dowstream decoder (reacting with a time-scale $\approx \frac{1}{q}$) be through summation of the depolarizations induced by the two spike trains. Finally, when additional spikes are present in one spike train, each spurious spike induces an increase in the total dissimilarity equal to half the maximal dissimilarity that a spike pair can reach. Similarly, an isolated spike induces a spurious depolarization of amplitude $\approx A$ once, while a maximally dissimilar spike pair reaches this depolarization twice (once for each spike of the pair).

Concerning the multi-unit spike train metrics, the different values of the between-neuron distinction degree k may be interpreted as different degrees of spatial separation (through reception by different neurons or by different parts of a dendritic tree) during the downstream combination of dACC signals (see Figure 3.1 (b)). Indeed, a maximal distinction degree can be implemented trough decoding by two different, unconnected neurons. Further, intermediate distinction degrees could (for instance) be implemented through different degrees of dendritic separation leading to tighter or looser requirements on the interspike interval to allow summation. In particular, threshold-like non-linearities in dendrites can prevent the summation of jittered EPSPs occuring in different dendrites (see Figure 3.1 (b, right)).

The metrics therefore accounts for plausible constraints of the downstream circuits in terms of signal processing, assuming the presence of one main decaying timescale for input filtering. Analysis techniques explicitly using exponential filtering for spike train classification were indeed found to behave almost identically to the Victor and Purpura distance [van Rossum (2001); Paiva et al. (2010); Chicharro et al. (2011)]. This is why the performance of the classification procedure is tantamount to the performance of these different decoding downstream circuits (rather than to the maximum amount of information that a perfect decoder, without any constraint, could reach).

Importantly, the presence of (task-epoch-specific) temporal structure does not necessarily cause an improvement of the decoding performance with an optimal value $q_{opt} > 0$ compared to $q = 0$. Indeed, temporal modulations may covary with spike-count differences, implying a redundancy between the spike-count based and spike-timing-based information. Further, the temporal information accessible to a biologically plausible decoder might reveal less robust than a time-integrated spike count. This is particularly likely to happen

in cases when the spike rate is consistently higher in one task-epoch compared to the other, leading to a between-task-epoch spike count difference that is consistent over time. This difference could be detected with more and more accuracy when evidence is accumulated over time through integration. This configuration (firing rate consistently higher in one task-epoch) seems to often qualitatively occur for dACC firing rates (see Figure 2.1 (c), Figure 4.1).

More precisely, this can correspond to cases when the downstream network needs to distinguish between two time-dependent (i.e., inhomogeneous) Poisson processes with firing intensities $\lambda_1(t)$ and $\lambda_2(t)$, such that for any time within the encoding window, $\lambda_1(t) > \lambda_2(t)$. The decoding of these two Poisson processes with the spike train metrics can be related to the estimation (from sample spike trains) of the dissimilarity between the two vectors of values of firing intensity $\vec{\lambda_1}$ and $\vec{\lambda_2}$ (i.e., the dissimilarity between two Post-Event Time Histograms PETHs, [Naud et al. (2011)]). We note that even though the spike train metrics decoding and the estimation of the dissimilarity between two PETHs are not exactly equivalent, both of them seek a processing mechanism permitting to distinguish well two different spiking processes. Hence, we will just use the estimation of the dissimilarity between PETHs in order to more formally illustrate the non-triviality of the improvement of such a distinction between spiking processes through temporal sensitivity of the processing mechanism. We will use the L1 norm of the difference between $\vec{\lambda_1}$ and $\vec{\lambda_2}$ as a measure of how much these vectors are dissimilar. Then, the decoding can be seen as an estimation of $\sum_t |\lambda_1(t) - \lambda_2(t)| = \sum_t (\lambda_1(t) - \lambda_2(t)) = (\sum_t \lambda_1(t)) - (\sum_t \lambda_2(t))$. As the sum of independent Poisson variables is also Poisson distributed, the decoding process is equivalent to the estimation of the difference of firing intensity of two Poisson variables, each of them being the temporal integration of one vector $\vec{\lambda_i}$. The minimum variance unbiased estimator of this difference actually is the difference of the means of two samples from the two variables [Bergamaschi et al. (2013)], i.e. this type of optimized estimator disregards spike timing. Hence, in this situation, temporal integration over the encoding window can permit maximal signal extraction despite ignoring temporal structure. This occurs through an averaging of samples which permit, in our case study, to average out the estimation errors over time. In contrast, a process which would compute absolute differences of spike count in small time bins would in this situation accumulate errors over different time bins, and would therefore be less efficient.

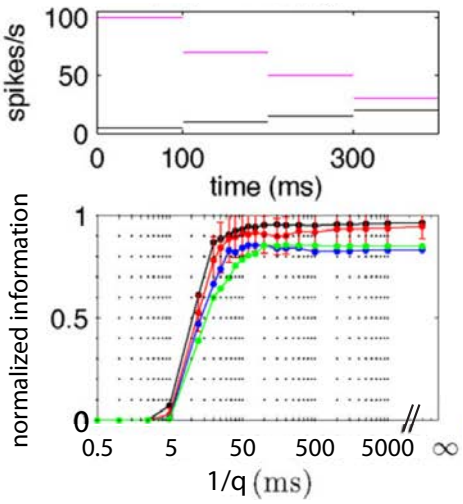


Figure 3.2: *Proof of principle for the non-triviality of the decoding improvement with temporal sensitivity.* Adapted from Fig. 2 of [Chicharro et al. (2011)]. The labeling of the lower graph was adapted to match our notation. *Top*: Rate profile of the two types of time-dependent Poisson processes producing the spike trains to be classified (each of the process is in a different color). *Bottom*: normalized information as a function of 1/q (in ms), for classifying data sets with 20 trials per stimulus. The data is the mean over 20 classifications, and the error bar is the standard deviation, shown for the red curve only. The different colors stand for different classification algorithms (all taking as a basis the VP distance; the red curve is exactly the algorithm that we used in Figure 4.3). Spike count classification (corresponding to q=0) does as well as, or better than, any temporal sensitivity, because the information in the temporal structure is redundant with the spike-count information.

Along those lines, previous articles reported an absence of timing-sensitivity-related information improvement even in the presence of category-specific temporal modulations in the spiking response [Oram et al. (2001); Chicharro et al. (2011)]. To illustrate, we reproduce here a figure from [Chicharro et al. (2011)] where this was the case (see Figure 3.2).

In conclusion, as pointed out in [Chicharro et al. (2011)], the spike-train-based classification does not detect all the existing timescales of the analyzed neuronal activity. Instead, the spike-train-based classification aims at testing whether the reliability of temporal structure could allow a plausible downstream decoder to take advantage of it, which relates to the biological plausibility of temporal information transmission [London et al. (2010)].

Algorithms and numerical methods

We ran all calculations on a cluster of 320 nodes (Consorzio Interuniversitario per le Applicazioni di Supercalcolo Per Università e Ricerca

CASPUR), on a private cluster (courtesy of S. Solinas) and on a PC laptop, using MATLAB (we adapted Victor's code, freely available at http://www-users.med.cornell.edu/~jdvicto/metricdf.html). For the single-unit decoding and response time analysis, we used Reich's c/MEX code and a modified MATLAB non-vectorized algorithm, respectively. For the multi-unit decoding analysis, we adapted Kreuz's vectorized algorithm in MATLAB code (to handle the case of empty spike trains). q was varied within $[0, 5, 10, 15, 20, 25, 30, 35, 40, 60, 80]s^{-1}$, whereas k was varied within [0,0.25,0.5,0.75,1,1.25,1.5,1.75,2].

3.2.2 Characterizing the nature of the informative spiking statistics

We used spike-time shuffling to investigate to what extent random samples from a time-varying trial-averaged rate density (as in Poisson neurons with time-varying rate) could underlie the advantage of the temporal structure for decoding [Victor and Purpura (1996)]. We will refer to this trial-averaged rate density as a Peri-Event-Time Histogram (PETH). For each cell and each task epoch separately, we grouped all spikes emitted in the interval $[0.001, 1]s$ post-feedback and randomly assigned each of them to a trial (repeated 1000 times, see Figure 4.10 (a) for a schematic explanation of the method). As a consequence, the number of spikes in each trial was actually drawn from a Binomial distribution with parameters $n = N_{spikes}$ and $p = \frac{1}{N_{trials}}$, which – following a common approximation – is close to a Poisson variable. Indeed, p was rather small (the trial number was usually large: 25^{th} quantiles were 14.25 and 51.25 for 1^{st} reward and repetition respectively); and the total number of spikes n was large (25^{th} quantiles were 53.75 and 175 for 1^{st} reward and repetition respectively). Under the Poisson approximation, spike counts restricted to sub-analysis windows are also Poisson (Raikov's theorem). This allowed us to build the spike-shuffled data for smaller analysis windows by simple pruning of the 1000 shuffled data of the largest window.

We used a second shuffling procedure to test to what extent information transmission could be determined by time-varying firing rates and spike-count variability as in the original data (see Figure 4.10 (d) for a schematic explanation of the method). In contrast to the previous shuffling method, this procedure considered that time-varying firing rate was modulated by a

multiplicative factor. This factor constrained the spike-count variability to fit the original data, and it was specific to each trial and time independent. Hence, this shuffling procedure not only conserved the PETH, but also the number of spikes present in each trial. To do so, for each cell, each task epoch, and each analysis window, we formed an ordered pool of all emitted spikes. Independently for each of these groups of spikes, we created 1000 shuffled data sets by randomly permuting the order of the spikes before reassigning to each trial the exact same number of spikes as in original data (without replacement).

Because both shuffling methods produced spike-shuffled data with the same number of trials as in the original data, the finite-sample information bias should be similar in both cases and should cancel when looking at the information difference, which was the relevant quantity. The bias was therefore not re-evaluated for this analysis.

3.3 Methodology for testing the negligibility of spike-sorting artifacts for the conclusions of the study

Spike-sorting relying on waveform shape (template) is reliable but does classify erroneously a small proportion of spikes. We explain below how we determined that these artifacts were unlikely to significantly affect our results.

- First, coincident spikes from different neurons will create 'mixture waveforms' that will be rejected by the algorithm (i.e. the spikes will not be assigned to putative neurons). Given that this phenomenon was very uncommon in our recordings, and and given that synchronized-spikes removal should decrease the reliability of both spike count and temporal coincidence decoding schemes, we do not expect this artifact to have a sizable impact on our analyses.
- Second, a small proportion of spikes accepted in a template are 'false positives' and belong to neurons different from the majority neuron. However, this is unlikely to favor spike-timing sensitive decoding over spike count decoding. Indeed, there was a bias toward having more cells firing prefentially during behavioral adaptation Quilodran et al. (2008). In

consequence, it was more likely that two randomly chosen neurons would show the same firing preference over task-epochs (i.e. either they would both fire more on average during the task-epoch requiring behavioral adaptation, or both fire more on average during the repetition task-epoch).

In this case, one could expect that the inclusion of the erroneously classified spikes would cause an increase in the reliability of both spike count and spike timing sensitive decoding (in the latter case, provided that the erroneously classified spikes are not damaging the reliability of the temporal pattern of activity of the majority neuron). We investigated this by comparing the decoding performance ($\langle I \rangle_t$, see Table 3.3) between different types of pairs of neurons when the decoder either ignored, or accounted for, the label of the spikes. More specifically, we used either spatially insensitive decoding (i.e. $k = 0$, a decoding which is insensitive to the neurons' identities), or spatially separate decoding ($k = 2$, a decoding which fully separates the activity of the two units). We contrasted the results between pairs of neurons for which both units fire preferentially in the same task-epoch, and for pairs with the two neurons firing preferentially in different task-epochs. A given neuron was said to fire preferentially in a given task epoch when its mean firing rate (in a [0.001,1]s window) was larger in this task epoch than in the other decoded task-epoch. We found that, when using spatially insensitive decoding (i.e. $k = 0$), pairs of neurons with the same firing preference performed better compared to pairs with different spiking preferences (pairs with significant coding, rank-sum test comparing: (i) spatially insensitive spike-count-based decoding $\langle I\left(q = 0, k = 0\right)\rangle_t$: $p_s < 10^{-4}$; (ii) spatially insensitive decoding with spike-timing sensitivity $\langle I\left(q_{opt}, k = 0\right)\rangle_t$: $p_s < 10^{-2}$). Note that we took $q_{opt} \approx 10s^{-1}$, as we will see later in the manuscript that this value of q appeared to maximize the mean information over neurons (see Figure 4.2). We also verified that the above-described difference between the pairs with same vs. different firing preference was not likely to reflect different intrinsic properties of the neurons between the two groups. Indeed, spatially separated decoding performed equivalently in the two groups ($\langle I\left(q = 0, k = 2\right)\rangle_t$ or $\langle I\left(q_{opt}, k = 2\right)\rangle_t$, all $p_s > 0.19$).

We now consider the (less probable) case when a spike-sorting error leads to the inclusion of erroneously classified spikes coming from a cell with a

different firing preference over task-epochs compared to the majority cell. Then, both spike count and timing-sensitive decoding are likely to be negatively impacted, given the small probability that the erroneously classified spikes can coincide with the precisely timed spikes of the majority neuron. Indeed, as we will see in Figure 4.12 (b-c), the temporal patterns of neuronal activity often differed between different neurons. Accordingly, when looking at pairs composed of two units with opposite firing preference, the information loss in spatially-insensitive ($k = 0$) decoding compared to spatially-separated ($k = 2$) decoding was not significantly distinct between timing sensitive and spike count codes (signed-rank test on $(\langle I\left(q=0,k=2\right)\rangle_t - \langle I\left(q=0,k=0\right)\rangle_t) - (\langle I\left(q_{opt},k=2\right)\rangle_t - \langle I\left(q_{opt},k=0\right)\rangle_t)$, all $p_s > 0.1$). This suggests that optimal timing-sensitive codes (i.e. $q = q_{opt}$) that are spatially-insensitive (i.e. the identity of the neuron which fires is unknown or ignored using $k = 0$) were not overall robustly better than spike-count at de-mixing two activities with opposite firing preference.

Overall, it is very unlikely that the 'false positive' spikes in a template, which are a minority and which do not appear to robustly favor spike-timing sensitive decoding, could sizably affect our results.

- Third, spikes of one neuron might pass from one template to another template (if the recording drifts), which could only potentially bias our pair of neurons analysis. The inter-electrode distance (150 μm of horizontal separation and, usually, different depths) made this phenomenon extremely unlikely between two different electrodes; this effect could only possibly affect pairs which templates were sorted on the same electrode. Such 'template exchange' might artificially produce low k_{opt} values in pairs recorded on the same electrode as compared to pairs recorded from different electrodes, and might artificially create the presence of pairs with $k_{opt} = 0$ (i.e. with the properties described in Figure 4.15).

  We tested this hypothesis by researching whether there was a consistent difference between pairs of neurons recorded from same vs. different electrodes. Note that such a difference may also arise if the inputs driving dACC are spatially segregated, making two closeby neurons more likely to receive similar inputs —as commonly observed, including in frontal areas [Schall et al. (1995)]. In this case, the differences between pairs recorded on the same vs. different electrodes could be specific to, say,

	1st reward discrimination			Errors discrimination		
	Monkey M	**Monkey P**	**Both monkeys**	**Monkey M**	**Monkey P**	**Both monkeys**
Distribution of k_{opt}	Diff. electrodes mean=1.19 median=1.5	Diff. electrodes mean=1.13 median=1.25	Diff. electrodes mean=1.15 median=1.25	Diff. electrodes mean=1.48 median=1.75	Diff. electrodes mean=1.17 median=1.25	Diff. electrodes mean=1.26 median=1.5
	Same electrode mean=1.21 median=1.375	Same electrode mean=1.16 median=1.25	Same electrode mean=1.18 median=1.25	Same electrode mean=1.08 median=1.125	Same electrode mean=0.87 median=1	Same electrode mean=0.96 median=1
	$p_{ranksum}$=0.95	$p_{ranksum}$=0.81	$p_{ranksum}$=0.83	$p_{ranksum}$=0.020	$p_{ranksum}$=0.012	$p_{ranksum}$=2.0 10^{-3}
Proportion of pairs for which k_{opt}=0	Diff. electrodes 4/43=0.093	Diff. electrodes 15/82=0.18	Diff. electrodes 19/125=0.15	Diff. electrodes 3/51=0.059	Diff. electrodes 16/128=0.125	Diff. electrodes 19/179=0.11
	Same electrode 4/20=0.20	Same electrode 4/32=0.125	Same electrode 8/52=0.15	Same electrode 7/34=0.21	Same electrode 13/48=0.27	Same electrode 20/82=0.24
	p_{fisher}=0.17	p_{fisher}=0.49	p_{fisher}=0.91	p_{fisher}=0.063	p_{fisher}=0.031	p_{fisher}=0.0037

Table 3.2: Comparison of the distribution of k_{opt} values, and of the proportion of pairs with $k_{opt} = 0$, between pairs recorded on different electrodes *vs.* the same electrode. Note that k_{opt} is the value of the parameter k that maximized time-averaged information (see [Table 3.3]). There were no significant differences for 1^{st}-reward discrimination, contrary to a consistent bias towards lower k values in the same electrode group expected if waveforms from different neurons were not well separated between different templates. The difference observed exclusively during errors classification most likely results from a spatial organization of the inputs responsible for the firing of dACC neurons during the error task-epoch. This can lead to more similar neural responses for closeby neurons as compared to more distant neurons.

errors discrimination, because the inputs driving the neurons at different moments of the task may have different spatial organization. In contrast, a generalized and consistent difference between these two groups may reveal either a bias due to spike-sorting or a generalized spatial structure of inputs.

Table 3.2 describes the results of:

- a rank sum test comparing distributions of k_{opt} values, where k_{opt} is the value of the parameter k that maximized time-averaged information (see [Table 3.3]).
- a Fisher test comparing the proportion of pairs with $k_{opt} = 0$

for significantly informative pairs recorded from the same vs. different electrodes.

For 1^{st} reward discrimination, the distributions of k_{opt} values and the proportion of pairs with $k_{opt} = 0$ were statistically identical among the pairs recorded from the same vs. different electrodes. By contrast, for errors discrimination the k_{opt} values were higher (and the proportion of $k_{opt} = 0$ smaller) for the group of pairs recorded from different electrodes. This result appears consistent with the existence of a spatial organization of inputs driving discharges during errors, and inconsistent with a (general) influence of spike sorting artifacts.

3.4 Methods for analyzing eye movements

We verified that purely motor differences between 1^{st} reward and repetition feedbacks were unlikely to produce the advantage of timing sensitivity for decoding. Here, we describe the methodology we used to perform this control analysis.

After target touch, arm-movements were largely a return from the target to the central 'lever' button occuring after gaze-shift. We therefore focused the analysis on eye movements, which were monitored with an infrared system (Iscan Inc., USA). We aimed at finding a threshold on the derivative of the recorded eye position which could define an eye movement. We filtered the signal with a gaussian of standard deviation 9 ms (changing this value by a few ms was not critical, see [Martinez-Conde et al. (2000)] for a similar approach). We then built a distribution of filtered eye-position derivatives, using peri-choice-saccade (0.1 s before to 0.5 s after targets onset) and post-reward (until +1 s) data, separately in X and Y. Distributions were gaussian-like supplemented with outliers (long tails). We used the threshold at which the data significantly differed from a gaussian — determined using the Grubbs Test implemented in the matlab file exchange function deleteoutliers [Shoelson (2003)] — to detect a movement in either X or Y. These X and Y thresholds matched well 'intuitive' saccade detection when we examined a large subset of traces. We actually chose the standard deviation of the filter for the position signal in order to maximize the gaussianity of the remaining distribution (after excluding outliers with the Grubbs Test).

Note that we did not differentiate between saccades and blinks (which both result in large derivative values of the recorded eye position), because they can trigger spiking in the same area [Bodis-Wollner et al. (1999)]. For simplicity, we

use the expression 'eye movement' to refer to any threshold crossing for recorded eye speed.

We characterized the eye motor activity between the go signal for target touch (occuring after target fixation) and 1s post-reward. Monkey P was very often breaking fixation before reward time (not shown), while monkey M was often maintaining fixation after reward time (as we will show in Figure 4.9 (a)). In both monkeys, differences could be seen between 1^{st} reward and repetition (e.g. in the number of saccades, latency of first saccade following the reward, as we will show in Figure 4.9 (a) for monkey M). Hence, there were differences in motor activity between 1^{st} reward and repetition task epoch. However, we note that these motor differences may actually not be reflected in the neuronal activity, or they may only impact neuronal activity indirectly, through a covariation with cognitive computations. Indeed, eye-movements may be correlated to attention and cognitive processing [Katnani and Gandhi (2013)]. This phenomenon seemed to occur at least for late eye-shifts in monkey M, as trials with late post-first-reward 1^{st} eye movement often led to a shorter response time of the monkey at the following trial (as we will show in Figure 4.9 (c)). Therefore, a correlation between these late saccades and neural activity would still be compatible with a cognitive correlate of the discharge.

In conclusion, we had to test whether purely motor differences between 1^{st} reward and repetition task epoch could participate to the advantage of temporal sensitivity for decoding. We will now describe how we determined that this was unlikely.

We focused on monkey M whose behavior allowed us to decode trials without any saccade or blink detected between the fixation period and the end of the analysis window (in Figure 4.9 (d,e,i)), or between the fixation period and 300 ms after the end of the analysis window (in Figure 4.9 (f,g,j)). This delay of 300 ms was chosen because it is likely to eliminate preparation activity directly triggering saccades (as the activity occuring, e.g., in the Frontal Eye Field [Hanes et al. (1995)]). We also excluded rare trials when, between saccade and reward time, the gaze had slowly drifted by more than one third of the inter-target difference. Because hand movements were almost always occuring after gaze shift, this process also minimized them. Beside, we stress that even though removing trials according to eye movements detection could induce some more pronounced differences in the proportion of the different targets between 1^{st} reward and repetition, this was very unlikely to favor purely motor-based classification, as target reach probably

happened too early (600 ms before the start of the analysis window) to still influence spiking.

Therefore, our trial-removal process would strongly reduce the advantage of temporal sensitivity for decoding if this advantage was reflecting motor activity (or premotor activity when the first movement occurs later than 300ms after the end of the analysis window). To test whether this was the case, we compared the improvement of information through temporal sensitivity ($I(q \approx q_{opt}) - I(q = 0)$) between the original data and data downsampled to remove putative motor or premotor activity. We also compared data downsampled to remove putative motor or premotor activity, to randomly downsampled data with identical trial number. Hence, the finite-sample bias should be similar between these two groups. Therefore, this bias should not impact the comparison between these two types of information values, and we indeed compared them directly without trying to evaluate the bias (in Figure 4.9 (e,g)). In addition, in order to consistently display bias-subtracted information in Figure 4.9 (d,f) as in all figures, the finite-sample information bias was evaluated as the mean information in 1000 shuffle data sets for which eye-movement free trials were randomly permuted between task-epochs.

Note that eye-movement data were only available in 38 significant neurons among the 61 from monkey M whose activity significantly distinguished between 1^{st} reward and repetition (i.e. those neurons used in Figure 4.2 (a, left)).

3.5 Methods for investigating the relation between neuronal activity and future behavior

For this analysis, only neurons with significant 1^{st} reward classification and with at least 5 available trials were used. Some additional analyses also tested different subgroups of this ensemble of neurons (in Figure 4.18 and Figure 4.19).

At the behavioral level, we focused on the response time which was defined as the time between the GO signal (for hand movement) following the 1^{st} reward, and the subsequent target touch. At the neuronal level, we aimed at quantifying how much a given spike train deviated from a (neuron-specific) prototypical 1^{st}

reward spike train.

3.5.1 Quantifying how much a spike train deviates from a prototype

For any given neuron, we wanted to quantify to what extent a spike train s was an outlier within the entire set of spike trains produced at 1^{st} reward, i.e. how much it deviated from the discharge 'typically' emitted during that epoch. To quantify this, it is possible to take the median of all pairwise dissimilarities between each spike train s and any other spike train s' emitted during the 1^{st} reward epoch. Indeed, in the space of neuronal responses, an outlier will be more dissimilar to the data set as a whole when compared to a data point that is close to the central point of the data set.

We now tackle the question of the choice of an appropriate dissimilarity measure.

The original Victor & Purpura distance $d(s, s')$ appears to not be optimal for this particular application. Indeed, it sums the costs to match any spike of train s to a spike of train s' (see subsection 3.2.1, [Victor and Purpura (1996)]). Thus, all pairwise distances involving a train with many spikes tend to be larger than those involving a train with little spikes. For instance, let $s = \{0.1, 0.5\}$ (i.e. it contains one spike at time $t = 0.1$ s and a second spike at $t = 0.5$ s) and $s' = \{0.11, 0.51\}$. Their distance is then $d_1(s, s') = 2 \cdot 0.01\, q$. Now, if $s = \{0.1\}$ and $s' = \{0.11, 0.51\}$, then $d_2(s, s') = 1 + 0.01\, q$. Therefore, if we take q to roughly match the temporal jitter of ± 0.01 s (i.e. $q = 100\ s^{-1}$), then $d_1 = d_2$, though during the first distance computation the spike matching was both as temporally precise as, and more reliable than, during the second distance computation. In order to avoid this scaling with spike number, we divided the Victor & Purpura distance by the number of times when two spikes (from the two trials) were 'coincident' (i.e., 'matched' during dissimilarity computation). Two spikes were considered 'coincident' when they were associated with a distance $d < (D_{max} = 2)$. There was no coincidence both in cases when a spike was deleted and then reinserted at the right time (for $q > 0$), and in cases when a spike was simply removed or added. Note that for $q = 0$, the number of 'coincidences' (i.e., 'spike matchings') is the spike count of the trial with less spikes. Therefore, the normalized distance can be expressed as:

$$d^*(s,s') = \frac{q}{N_c} \cdot \sum_{i}^{N_c} \left| t_s^i - t_{s'}^i \right| + \frac{C}{N_c} = q \cdot \langle dt \rangle + \frac{C}{N_c} \tag{3.3}$$

where N_c denotes the number of coincident spike pairs, t_s^i the time of the i^{th} coincident spike in train s, $\langle dt \rangle$ the mean jitter among coincident spikes, and C the total cost for inserting and/or deleting spikes. The first term quantifies the dissimilarity due to coincident (i.e., 'matched') spikes, whereas the second one is the dissimilarity due to unmatched spikes. For $q > 0$, this measure quantifies the reliability of temporal coincidence detection between two spike trains. For $q = 0s^{-1}$, it quantifies the absolute spike count difference relative to the shared spike count. In both cases, the normalized distance behaves similarly to an inverted signal-to-noise ratio. In this interpretation, the signal is taken as the coincident spikes. The noise is the unmatched spikes, and the temporal jitter of coincident spikes relative to the considered 'coincidence window' for $q > 0$.

In the absence of coincident spikes, we simply used the original Victor & Purpura distance. For $q = 0$ s^{-1}, the absence of coincident spikes only happens when one spike train is empty. In this case, some intuitive order relations are conserved. Let s_x denote a spike train containing x spikes. Then: $d(s_0, s_x) > d(s_0, s_y)$ iff $x > y$, and $d(s_0, s_x) > d(s_1, s_x)$ if $x > 1$. For $q > 0$, the absence of matching spike could also happen when the distance between two spike trains s_x, s_y is maximum and equal to $x + y$, because no spikes are close enough in time to be advantageously matched. In this case, the distance grows with the number of spikes that are unmatchable, i.e. very dissimilar, which appears sound. We note that our results show (as will be visible soon) an increase of information driven by temporal spike matching, which implies that this 'no coincident spikes' situation was likely to be unfrequent.

Note that the new distance we designed has a different purpose and effect from the previously proposed division by the sum of spike count in the two spike trains and, more generally, from other re-scaled spike train dissimilarity measures [Naud et al. (2011)]. Indeed, rather than bounding the measure, or merely averaging some jitter statistics, we tried to build a measure that would evaluate dissimilarities between spike trains as perceived by different plausible decoders which are more or less sensitive to spike timing and spike count, without being biased by the number of spikes. Notably, we did not want the spikes that could not be matched to enter in the normalization factor for the dissimilarity measure (which would happen with a simple division by spike

count).

Finally, we stress that as expected, the normalized distance $d^*(s, s')$ showed similar classification ability as compared to the classical Victor & Purpura distance d. Indeed, for any spike train s, since both the intra- and inter-task epoch distances d to s will increase with the spike count of s, a smaller d for a given task epoch still indicates a greater similarity relative to the other task epoch(s). To corroborate this hypothesis, we tested the 1^{st} reward classification with the normalized metrics. To do so, we used the very same trials that have been extracted for the response time analysis. Both the number and the identity of the significant neurons were consistent with those found with the classical metrics (Monkey M: 65 significant neurons vs. 61, of which 57 are shared; monkey P: 50 significant neurons in both cases, 44 shared). The classification results were also equivalent, as confirmed by a rank sum test comparing the maximum (over timing sensitivity values) time-averaged information among significant neurons (all $p_s > 0.74$). In addition, the normalized metrics uncovered an increase of time-averaged information with timing sensitivity adaptation, independently in both monkeys (Friedman ANOVA on time-averaged information $< I >_t$, all $p < 10^{-8}$; $q_{opt} = 15s^{-1}$ and $10s^{-1}$ for monkey M and P respectively showed higher rank than $q = 0$ after post-hoc comparisons with Tukey's honestly significant criterion).

3.5.2 Testing whether deviation from prototype is predictive of response time

Let $\widetilde{r}$ denote the median value of observed response times, T_+ be the set of 1^{st} reward trials followed by a response time larger than $\widetilde{r}$, and T_- the set of trials followed by a response time lower than $\widetilde{r}$. For each spike train s, we calculated the dissimilarity between s and prototypical 1^{st} reward activity (i.e. $median\left(d_q^*(s, s')\right)_{s' \in 1^{st}\ reward,\ s' \neq s}$, similar to the spike train classification analysis). We then defined $\overline{D_{T_+}}$ ($\overline{D_{T-}}$) as the mean over all $s \in T_+$ (T_-) of the dissimilarity between s and prototypical 1^{st} reward activity. We finally computed the overall difference of deviation from the prototypical discharge at 1^{st} reward as $\overline{D} = \overline{D_{T_+}} - \overline{D_{T_-}}$. $\overline{D}$ was computed for multiple time window lengths: from 100 ms to 1 s post-feedback time, by increments of 100 ms. Finally, a bias score $b = \sum_{-\ bias\ win} log\left(p_{signed\ rank}\left(\overline{D}\right)\right) - \sum_{+\ bias\ win} log\left(p_{signed\ rank}\left(\overline{D}\right)\right)$ was

computed. To determine which analysis windows were positively or negatively biased, we used the signed-rank statistics, which relies on the ranking of the $abs\left(\overline{D}\right)$ values (where $abs\,(\cdot)$ is the absolute value). Therefore, $+\ bias\ win$ is the set of "positive bias windows" which contains those analysis windows for which the sum of these ranks for positive values of $\overline{D}$ was larger than the sum of these ranks for negative values of $\overline{D}$. Similarly, $-\ bias\ win$ is the set of "negative bias windows" for which the sum of ranks for negative values was larger than the sum of ranks for positive values. A positive (negative) bias in a given window would cause a corresponding increase (decrease) in b. To assess the significance of the bias score b, 1000 surrogate data sets, in which the difference between high and low response time groups was eliminated, were compared to the real data. For each surrogate, and independently for each neuron, the sign of all $\overline{D}$ values (for all analysis windows) had a 0.5 probability to be changed. The p-value was computed as the proportion of surrogate data sets leading to higher or equal $abs\,(b)$ as the real data.

A similar analysis was done to test whether firing rates could also relate to response time. To do so, $\overline{D}$ was replaced by the difference in mean firing rate between high and low response time trials, $\overline{D_{rate}}$.

3.5.3 Testing whether the prediction of behavior from neuronal activity is different between $q = 0$ and $q \approx q_{opt}$

The temporal sensitivity q leading to best 1^{st} reward discrimination in the population ($q \approx q_{opt}$) and q=0 were compared (signed-rank test). To do so, we computed the mean values of $\overline{D}$ over analysis windows from 100 ms to 1 s post-feedback time, by increments of 100 ms. Similar results were found when assessing the optimal q value by using either the original Victor & Purpura distance d_q (in Figure 4.17), or the normalized distance d_q^* (in Figure 4.18).

For one monkey (monkey P), we will see later in the manuscript that this test was not significant. This could reflect either a negligible impact of spike timing sensitivity on the $\bar{D}$ measure, or the fact that $d^*(q_{opt})$ and $d^*(q = 0)$ were yielding equally strong, but still rather different neuronal-behavior correlations. For instance, $d^*(q_{opt})$ and $d^*(q = 0)$ could lead to large $\bar{D}$ values in different neurons and/or for different analysis window durations. Hence, we designed a

method to investigate this question. We noticed that if temporal sensitivity had a negligible impact on the $\bar{D}$ measure, then the difference $Diff_{\bar{D}} = \bar{D}(q_{opt}) - \bar{D}(q = 0/s)$ should be negligible noise. Under this (null) hypothesis, adding to $\bar{D}(q = 0/s)$ a surrogate noise $Diff_{\bar{D}}^{surr}$ – with statistics that are similar to those of $Diff_{\bar{D}}$ – should lead to a surrogate of $\bar{D}(q_{opt})$: $Diff_{\bar{D}}^{surr} + \bar{D}(q = 0/s) = \bar{D}(q_{opt})^{surr}$. This $\bar{D}(q_{opt})^{surr}$ should then have a similar bias score b to the bias score of the original $\bar{D}(q_{opt})$. We tested this hypothesis, creating 1000 surrogates $Diff_{\bar{D}}^{surr}$ from $Diff_{\bar{D}}$ by randomly shuffling the values of $Diff_{\bar{D}}$ between neurons (identical conclusions were also reached when shuffling between analysis windows or between both neurons and analysis windows). Importantly, only 2 % of these surrogates had bias scores b superior or equal to the one of the original $\bar{D}(q_{opt})$ (using analysis windows ending between 0.1 and 1s by steps of 0.1 s for bias score computation). In other words, the null hypothesis (stating that temporal sensitivity at $q_{opt} = 10/s$ was only producing spurious negligible changes in d^* relative to $q = 0$) could be rejected with a p-value of 0.02. Additionally, we would like to briefly mention that we also implemented a similar test while computing the bias score using only analysis windows during which $q = 0$ was leading to a substantial value of $\bar{D}$ (see Figure 4.17, analysis windows between 250 and 450 ms, increasing in steps of 50 ms). This gave a similar result (p=0.011), strengthening our conclusion that the effects seen when using $d^*(q_{opt})$ vs. $d^*(q = 0)$ were at least partially separate.

3.6 General statistics

Time-averaged information $\equiv \left\langle I \right\rangle_t$

Difference in mean spike count between task epochs $\equiv \left\langle N_{adapt} \right\rangle_{trials} - \left\langle N_{repet} \right\rangle_{trials}$

Normalized absolute difference in mean spike count between task epochs $\equiv \frac{\left| \left\langle N_{adapt} \right\rangle_{trials} - \left\langle N_{repet} \right\rangle_{trials} \right|}{\left\langle N_{adapt} \right\rangle_{trials} + \left\langle N_{repet} \right\rangle_{trials}}$

Optimal timing sensitivity $q_{opt} \equiv mean\left(\underset{q}{\operatorname{argmax}} \left(\left\langle I \right\rangle_t \right) \right)$

Optimal distinction degree between units $k_{opt} \equiv mean\left(\underset{k}{\operatorname{argmax}} \left(\left\langle I \right\rangle_t \right) \right)$

Temporal structure related gain of information $\equiv \frac{\max_q \left(\left\langle I \right\rangle_t \right) - \left\langle I_{q=0} \right\rangle_t}{\left\langle I_{q=0} \right\rangle_t}$

Fano factor estimate $F \equiv \frac{\widehat{var}(C_t)}{\langle C_t \rangle} = \frac{\frac{\sum_{i=1}^{n_{trials}} \left(C_t^i - \langle C_t \rangle \right)^2}{(n_{trials}-1)}}{\langle C_t \rangle}$, where C_t is the random variable counting the number of spikes fired by a neuron in a given analysis windows during one task-epoch, and C_t^i is its realization in a given trial among the n_{trials} available trials.

Gain in the pair relative to the best single unit $\equiv \frac{\max_{q,k} \left(\left\langle I^{pair} \right\rangle_t \right) - \max_{cells,q} \left(\left\langle I^{single} \right\rangle_t \right)}{\max\left(\max_{q,k} \left(\left\langle I^{pair} \right\rangle_t \right), \max_{cells,q} \left(\left\langle I^{single} \right\rangle_t \right) \right)}$

Information imbalance between two units $\equiv \frac{\left| \max_q \left(\left\langle I^{cell1} \right\rangle_t \right) - \max_q \left(\left\langle I^{cell2} \right\rangle_t \right) \right|}{\max\left(\max_q \left(\left\langle I^{cell1} \right\rangle_t \right), \max_q \left(\left\langle I^{cell2} \right\rangle_t \right) \right)}$

For pairs with $k_{opt} = 0$:

Information gain when not distinguishing between neurons $\equiv \frac{\max_q \left(\left\langle I^{pair}_{k=k_{opt}=0} \right\rangle_t \right) - \max_q \left(\left\langle I^{pair}_{k=k\max=2} \right\rangle_t \right)}{\max_q \left(\left\langle I^{pair}_{k=k_{opt}=0} \right\rangle_t \right)}$

Between-neuron spike coincidence index $\equiv \max\left(P^{intra}_{adapt}, P^{intra}_{repet} \right) - P^{inter}$

with P denoting the proportion of trials for which between-neuron spike-matching(s) did impact the Victor & Purpura distance $d_{q_{opt}, k_{opt}}$, for the analysis window that maximizes I^{pair}.

Table 3.3: The angle brackets denote averaging; t denotes time average over the ensemble of analysis windows beginning 1 ms after the feedback and ending from 100 ms to 1 s (by steps of 100 ms). Information values I were always normalized and bias corrected unless mentioned. We therefore simply refer to them as "information" throughout the text. "adapt" stands for behavioral adaptation task epochs (either errors or 1^{st} reward); "repet" stands for repetition task epochs. N is the spike count in a window between 1 and 1000 ms after feedback onset. $argmax_y$ $(f(y))$ is the point y_o of the argument y for which the function f attains its maximum value.

Table 3.3 summarizes an additional set of employed statistical measures. The latter were often non-normal, therefore non-parametric tests were considered ($p \leq 0.05$ was considered as statistically significant):

- correlations were assessed using Spearman coefficient with a permutation test (or a large sample approximation)
- distributions were compared with the 2-sided Kolmogorov-Smirnov test
- central tendencies were compared between 2 unpaired (resp. paired) distributions with the 2-sided ranked-sum (signed-rank) test
- deviation of distributions from 0-centered-symmetry was also tested with the 2-sided signed-rank test

When testing pairs of units, one limitation was that some pairs happened to share a neuron, and hence were correlated (in particular if non-shared neurons were discharging significantly less than the shared one). This was problematic for analyzing the optimal temporal sensitivity, which is not a parameter accounting for the interaction between neurons, and which can be impacted more by the neuron which fires the most. We therefore verified that the significance of the advantage of the temporal sensitivity during paired decoding could be reached without overlapping pairs (positivity of $max_k\left(\langle I\left(q=10s^{-1}\right)\rangle_t\right) - max_k\left(\langle I\left(q=10s^{-1}\right)\rangle_t\right)$, signed-rank test, $p \leq 0.05$ in 1000/1000 random down-samplings to non-overlapping pairs). Note that, in contrast, interaction parameters such as the information gain or k_{opt} are truly pair specific, implying that it was reasonable to keep overlapping pairs for the analysis. Note that although most statistical tests we present were carried out by pooling data from both monkeys, consistent trends were observed for both individuals.

The standard error of the mean for the variable X was taken as $\frac{\sqrt{\frac{\sum_{i=1}^{N}(X_i-\langle X\rangle)^2}{N-1}}}{\sqrt{N}}$. Error bars for the median were taken as $\pm\frac{interquartile\ range}{1.075\ \sqrt{n}}$, as scaling the median with this standard-error-like value approximately gives a t distribution with (n-1) degrees of freedom [Hoaglin et al. (1983)]. This therefore is approximately a 70% two-sided confidence interval ($t_{0.7,(n-1)} \approx 1$ for a large range of n values, and the confidence interval is $t_{0.7,(n-1)}$ times the standard error).

Unless mentioned otherwise, the boxplots represent the median at the notch,

the 25^{th} and 75^{th} quantiles as horizontal lines, and the whiskers extend until at most 1.5 times the interquartile range beyond the closest quartile. Finally, beyond these whiskers, outliers are indicated as red crosses (unless mentioned otherwise).

Chapter 4

Testing decoders of the behavioral adaptation signals emitted by dorsal Anterior Cingulate Cortex neurons

To investigate temporal coding in dACC, we analyzed the activity of 145 and 189 individual neurons from monkey M and P, respectively.

4.1 Optimal temporal sensitivity improves decoding of single units' behavioral adaptation signals

We first tested how single-trial single-unit dACC activity could send signals that could drive behavioral adaptation after feedback. Behavioral adaptation occurred either after the 1^{st} reward (thus switching from exploration to repetition) or after any error during exploration (see Figure 2.1 a). Signaling the need for adaptation requires that spike trains emitted during either 1^{st} reward or errors can be discriminated from those emitted during repetitions (referred to as 1^{st} reward and error discrimination analyses, respectively). Neurons in dACC showed early post-feedback responses specific to behavioral adaptation [Quilodran et al. (2008)]. Therefore, we analyzed spike trains starting at the onset of the feedback delivered 600 ms after target touch. We will refer to any post-feedback time interval (i.e. following either an error, or 1^{st} reward, or repetition) as a task epoch. We quantified to what extent spike trains emitted during different task epochs were discriminable by a downstream decoder by classifying them based

on a spike train dissimilarity measure [Victor and Purpura (1996)]. We briefly remind the reader that this dissimilarity measure computed the minimal cost to transform the first spike train into the second one through two possible recursive operations: (i) adding or removing a spike, for a cost of 1; and (ii) changing the timing of a spike by dt, for a cost of $q\ dt \leq 2$. Note that the maximum cost allowing two spikes to be temporally matched (coincidence detection) is 2 because it corresponds to the total cost of removing one spike and adding another spike at any desired time (see Figure 3.1 a and section 3.2). This measure allows different temporal sensitivities of a downstream decoder to be evaluated by varying the parameter q. A value of $q = 0s^{-1}$ describes a decoder sensitive to pure spike count. On the other hand, a larger q value corresponds to a decoder sensitive to precise spike times. The larger the q value, the smaller the maximum interspike interval leading to coincidence detection, and the more the decoder disregards spike count. We stress again that even when the neural activity is temporally structured, sensitivity to spike timing does not necessarily improve decoding. For instance, spike timing and spike count might provide redundant information and then a neural integrator could be more robust (see section 3.2.1).

We quantified the classification performance (i.e. how well, on average, a spike train was correctly associated to the task epoch with the most similar activity) by computing the mutual information between the predicted distribution of spike trains across task epochs and the true distribution (see section 3.2.1). Throughout this thesis, mutual information values are expressed as percentage of the maximum value corresponding to perfect discrimination. Information values were computed for different analysis windows, all starting 1 ms after feedback time and with increasing duration. In this way, the state of a putative decoder of dACC feedback-related discharges could be evaluated at different delays after the start of the decoding process.

Finally, we stress that unless mentioned otherwise, we display results among all "significant" neurons. We remind the reader that these neurons emitted spike trains that could be classified between task epochs with a higher accuracy than chance level (permutation test, $p < 0.05$; see section 3.2.1).

4.1.1 Optimal temporal sensitivity mediates information improvement in a majority of single neurons

Consistent with previous results focusing on spike count only [Quilodran et al. (2008)], we found that most dACC neurons with significantly selective task-epoch activity fired more during behavioral adaptation periods (i.e. post 1^{st} reward and/or error feedbacks) compared to reward in repetition (see Figure 4.1, and Figure 4.8 (a)).

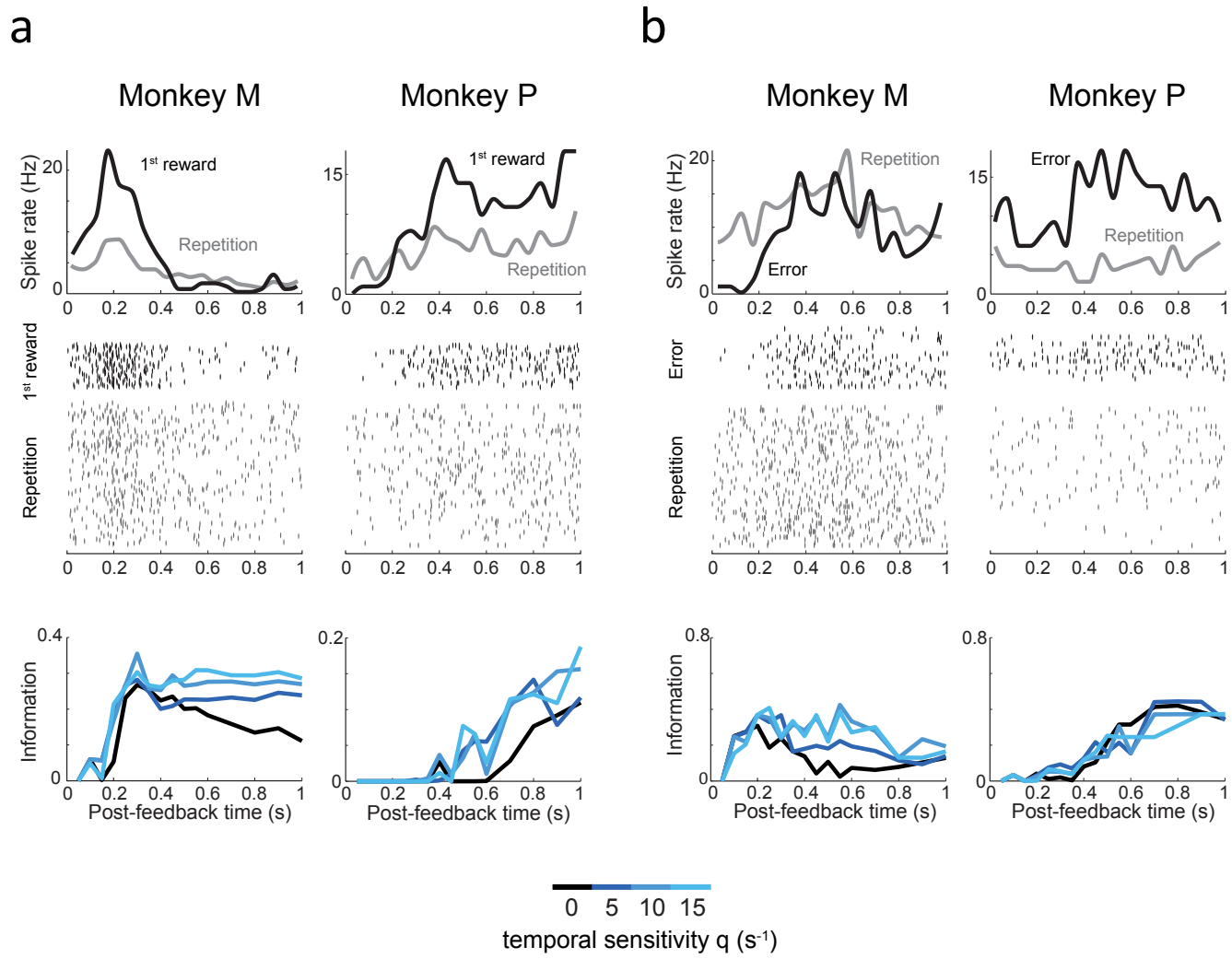


Figure 4.1: *Examples of single-unit dACC activities decoded with different temporal sensitivities.* **(a)** Spike densities (top) and raster plots (middle) during 1^{st} reward (black curve) and repetition (grey curve) task epochs. The classification performance between 1^{st} reward and repetition spike trains (i.e. information) is shown in the bottom graphs, the time in the abscissa being the time at which the analysis window (and thus, the decoding process) ends. Two neurons, from the two monkeys, are shown. These samples show that temporal sensitivity can improve classification performance. **(b)** Same as (a) but for errors and repetition in two other neurons from the two monkeys.

We tested whether temporal sensitivity would consistently tend to improve information transmission among all these significant neurons, compared to spike count. Importantly, for most neurons, timing-sensitive decoding of spike trains ($q>0$) conveyed more information than spike count ($q=0$; Figure 4.2).

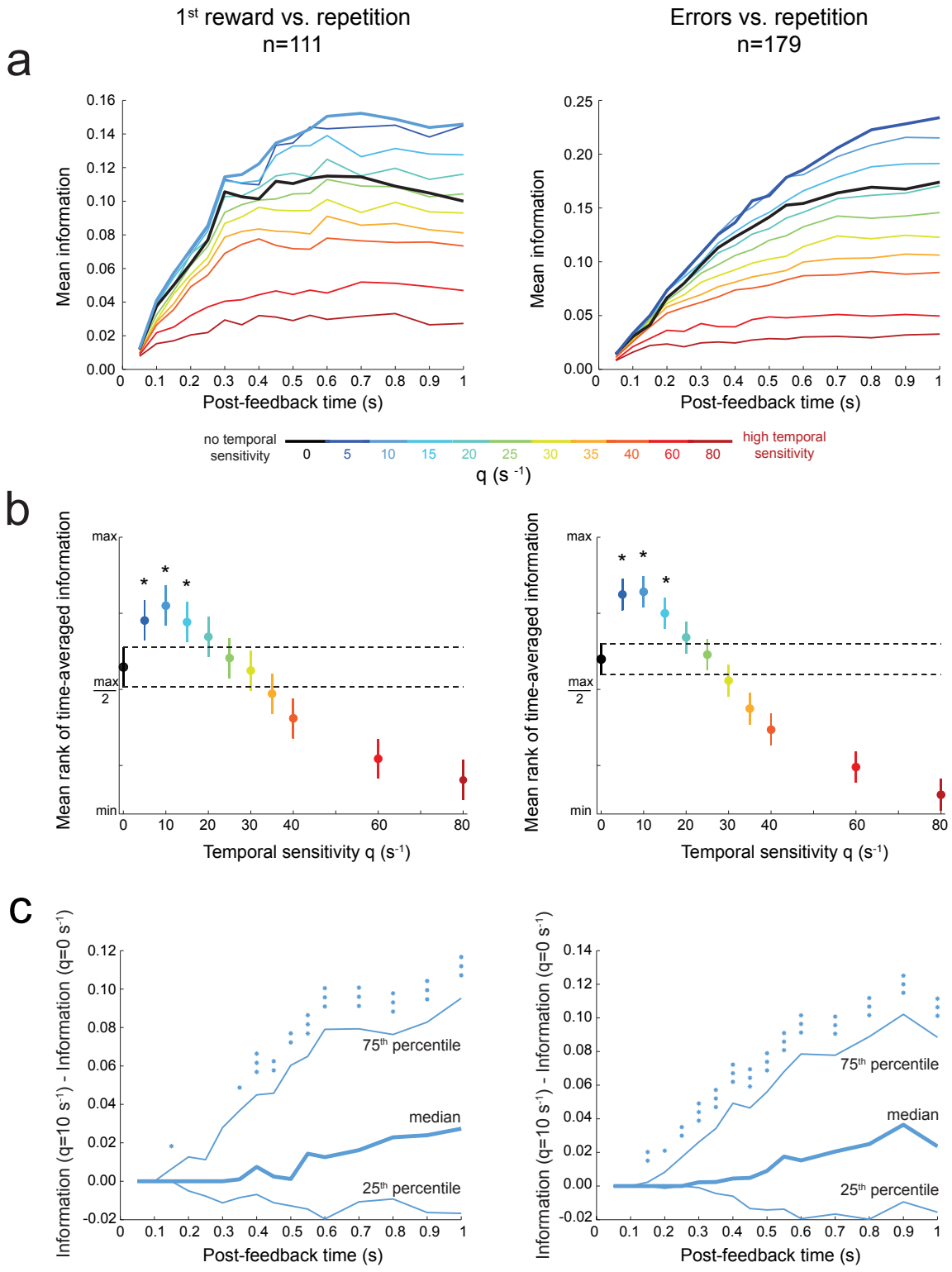


Figure 4.2: *Optimal temporal sensitivity improves decoding of single unit behavioral adaptation signals.* **(a)** Time course of the mean information (averaged among significant cells) as a function of the decoding temporal sensitivity (q). Information values were computed over increasing post-feedback time windows (ending at the time indicated by the x-axis). *Left*: Discrimination between 1^{st} reward and repetition task epochs. *Right*: Discrimination between error and repetition task epochs. **(b)** Time-averaged information $\langle I \rangle_t$ (see definition in Table 3.3) for different temporal sensitivities (q). The ordinate axis is the normalized mean rank of $\langle I \rangle_t$. After a Friedman test, post-hoc comparisons with Tukey's honestly significant difference correction were used for the 95% confidence intervals. Temporal sensitivities q>0 that were performing significantly better compared to $q = 0$ are indicated by a star. **(c)** Distribution of the difference of information between optimal temporal decoding ($q_{opt} \approx 10s^{-1}$) and spike-count decoding ($q = 0s^{-1}$). Stars indicate the significance of signed-rank tests (the null hypothesis is the symmetric distribution around zero): *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$.

We characterized this effect by looking at the time course of information (averaged across neurons with significant decoding), for different q_s (see Figure 4.2 (a)). For each value of q, the information increased as post-feedback spiking accumulated with time. Temporal sensitivity influenced both the maximum amount of information and the speed at which it increased. Importantly, adapted temporal sensitivity provided a sizable gain (15-40%) in mean information compared to spike count. Values of q within [5,10,15]s^{-1} led to a significant increase in time-averaged information $\langle I \rangle_t$ (defined in the caption of Table 3.3; Figure 4.2 (b); Friedman ANOVA, global effect on all considered q values: $p<0.001$). This effect was robust early after the feedback and for all subsequent times (Figure 4.2 (c)).

The same phenomena were observed when using a classifier that was biased towards nearest neighbors, instead of the unbiased classification that we use for all other figures of the dissertation (see section 3.2.1, Figure 4.3). The nearest-neighbors biased classification was actually less robust (leading to less significant neurons), and that is why we display results using the unbiased classification. However, we verified that the main results would hold for both classification techniques. Actually, we will show (in subsection 4.2.1) that the spike count variability was large in our data. This suggests that any classifier biased towards some type of outliers would probably perform worse than the classifier we used, and would impact spike-count decoding more negatively compared to temporally sensitive decoding.

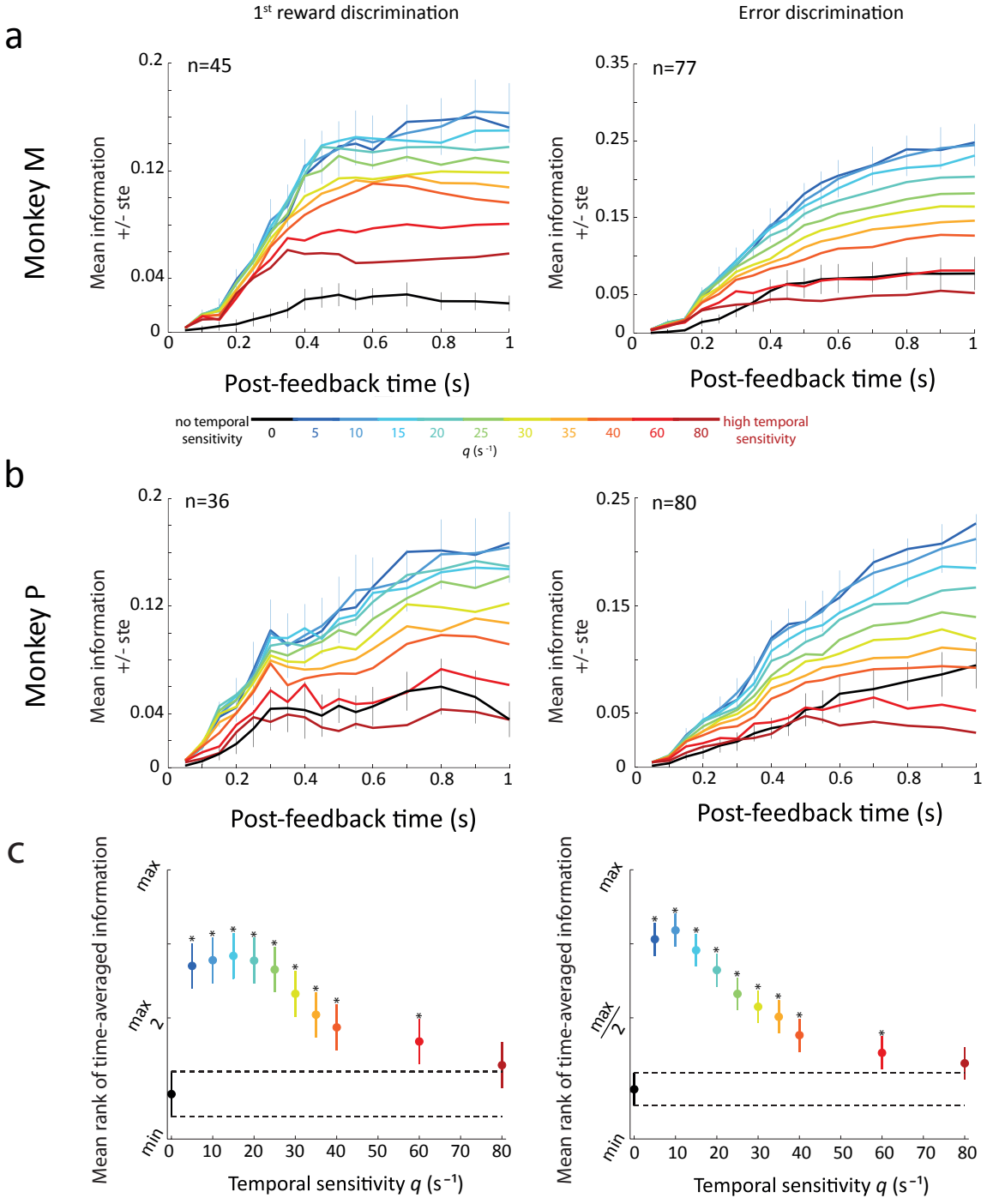


Figure 4.3: *Information gain through temporal sensitivity using a classification biased toward closer neighbors instead of the unbiased classification.* Information gain through temporal sensitivity was also observed when the classification of spike trains was biased toward smaller dissimilarities rather than determined by the median dissimilarity to spike trains of a task-epoch (see section 3.2.1). Results in this figure are for the neurons with significant discrimination ability (permutation test, see section 3.2.1); note that the number of significant units is smaller than with the classification method using the median (see Figure 4.2). **(a,b)** show the time course of the mean information (over neurons) for 1^{st} reward (left) and errors (right) discrimination, as a function of timing sensitivity q, separately for the two monkeys. **(c)** Results of the post-hoc comparisons (with Tukey's honestly significant criterion correction for multiple comparison) of a Friedman ANOVA comparing the time-averaged information $\langle I \rangle_t$ between temporal sensitivities. Note that the slight differences in the rankings of q-values between **(a,b)** and **(c)** are due to the fact that the mean over neurons is more sensitive to outliers with high values, while the average rank is determined by the consistency (over the population of single units) of the within-neuron rankings of $\langle I \rangle_t$ between different q-values.

Also, the advantage of temporal decoding over spike count decoding was robust in both monkeys individually (Figure 4.4).

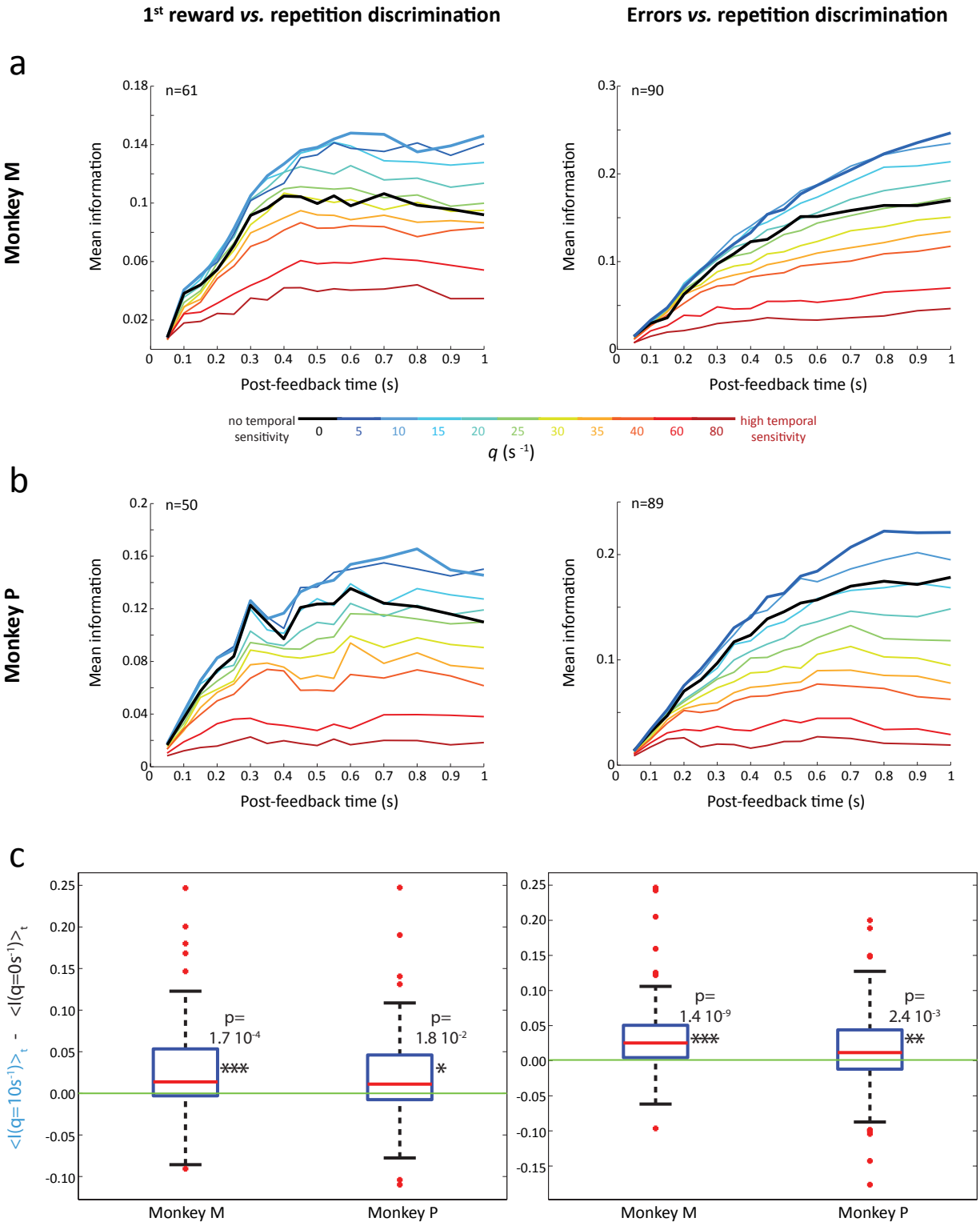
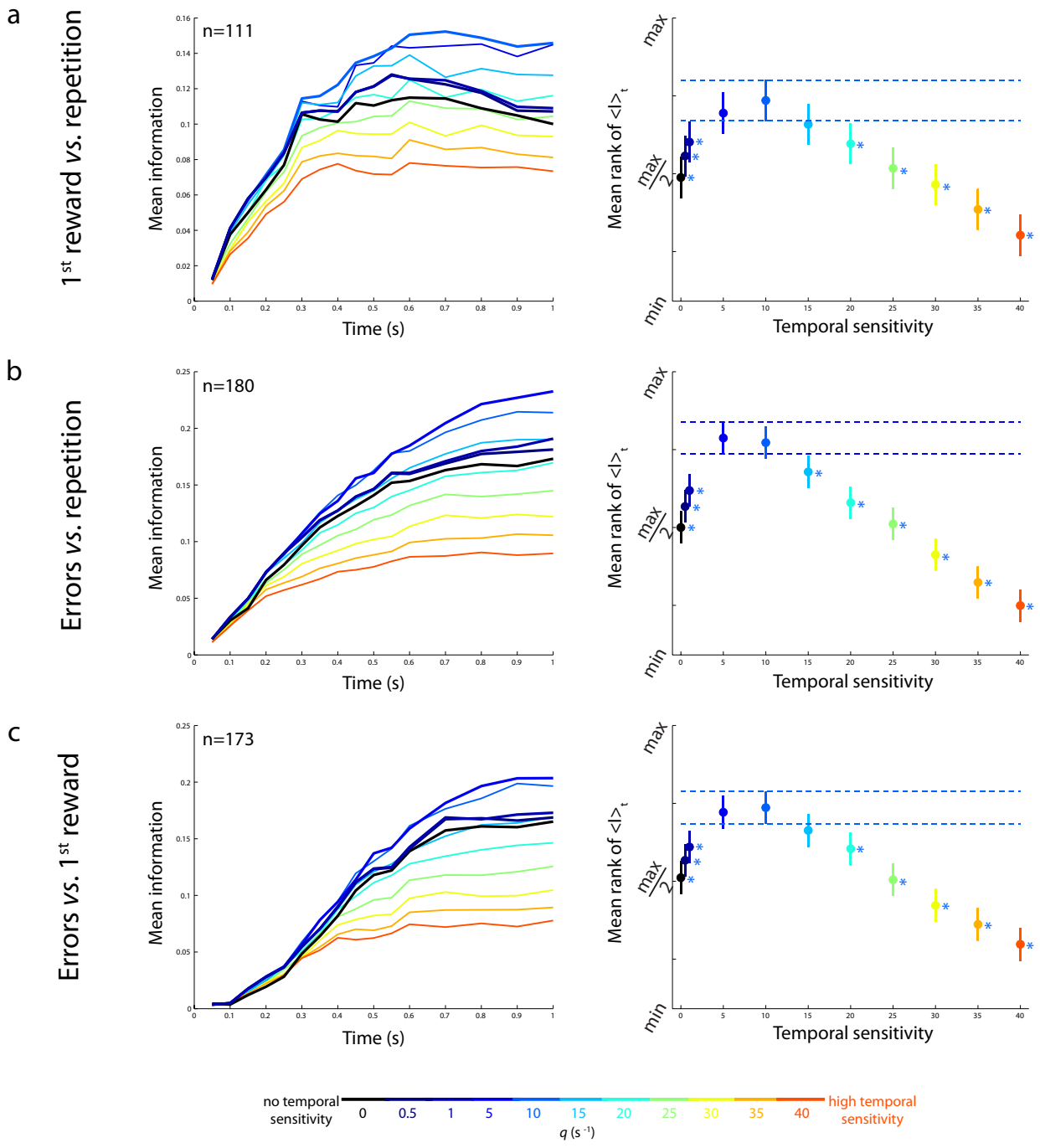


Figure 4.4: *Robustness of spike-timing information in both monkeys.* The improvement of decoding trough spike-timing sensitivity was robust in both monkeys. The left part of the figure describes the result of the discrimination between 1^{st} reward and repetition, and the right part describes errors *vs.* repetition discrimination. **(a,b)** Time-course of the mean information over neurons, for different temporal sensitivities of the decoder q as indicated on the color scale, for monkey M and P respectively. **(c)** Difference of time-averaged information $\langle I \rangle_t$ between temporal decoding ($q_{opt} \approx 10s^{-1}$) and spike-count decoding ($q = 0s^{-1}$). The p-value of a signed-rank test indicates that in both monkeys individually, temporal sensitivity induced a robust increase of information (all $p_s \leq 0.018$). The horizontal green line marks the 0 value.

In addition, we verified that very small temporal sensitivities (compatible with an imperfect integrator implementing a slowly decaying memory: $\tau \geq 1s$), were leading to significantly less information than optimal temporal decoding (Figure 4.5 (a,b)). Given that the behavioral task required to maintain the memory of the adapted behavioral strategy over several seconds, this suggests that a slow enough leaky integrator would indeed read out a less robust signal from dACC spike trains than a timing-sensitive decoder (Figure 2.1).


a
1st reward vs. repetition
n=111
Mean information
Time (s)
Mean rank of <I>t
max
max/2
min
Temporal sensitivity
b
Errors vs. repetition
n=180
Mean information
Time (s)
Mean rank of <I>t
max
max/2
min
Temporal sensitivity
c
Errors vs. 1st reward
n=173
Mean information
Time (s)
Mean rank of <I>t
max
max/2
min
Temporal sensitivity
no temporal sensitivity
0 0.5 1 5 10 15 20 25 30 35 40
high temporal sensitivity
q (s-1)

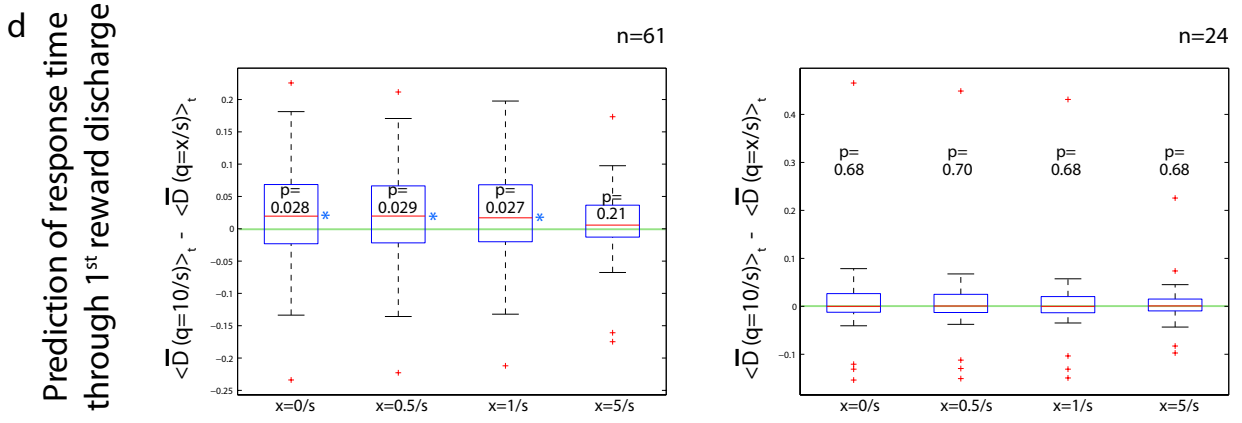

d
Prediction of response time through 1st reward discharge
n=61
n=24
<D (q=10/s)>t - <D (q=x/s)>t
p= 0.028
p= 0.029
p= 0.027
p= 0.21
p= 0.68
p= 0.70
p= 0.68
p= 0.68
x=0/s
x=0.5/s
x=1/s
x=5/s

Figure 4.5 ***(previous page)*:** *Using a small temporal sensitivity (compatible with decoding by an imperfect integrator) leads to identical conclusions to using q=0/s (perfect integration) in single units.* We test: (i) $q = 0.5s^{-1}$, approximately equivalent to an exponential leak time-scale $\tau = \frac{1}{q} = 2s$ (see Figure 2.1 and Figure 3.1 (a), and section 3.2.1). This is the minimal time-scale for a downstream leaky neuronal integrator which has to hold in memory the behavioral adaptation signals (and/or the behavioral strategy signals) for up to 3-6 s as required during the task (in case of fixation break). (ii) $q = 1s^{-1}$, approximately equivalent to an exponentially decaying time-scale $\tau = 1s$, as a more stringent test. **(a,b,c)** Classifying spike trains: 1^{st} reward *vs.* repetition **(a)**, errors *vs.* repetition **(b)**, errors *vs.* 1^{st} reward **(c)**. We used neurons reaching significant classification with any q-value (including $q = 0.5$ and $1s^{-1}$, permutation test, Methods), leading to only one more significant neuron compared to Figure 4.2 (for errors *vs.* repetition classification, monkey P). Left: time-course of the mean information over neurons. Right: results of post-hoc comparisons of the time-averaged information $\langle I \rangle_t$ after a Friedman anova, using the Tukey's honestly significant criterion correction. Q-values with significantly smaller performance than q_{opt} are marked by a star. In all considered cases, both $q = 0.5s^{-1}$ and $q = 1s^{-1}$ were leading to significantly smaller $\langle I \rangle_t$ than q_{opt}. In both monkeys individually, $q = 0.5s^{-1}$ and $q = 1s^{-1}$ had (at least qualitatively) lower average rank than $q = 10s^{-1}$ and $q = 5s^{-1}$. The Friedman test was restricted to $q \leq 40s^{-1}$, focusing on q-values for which classification was not too noisy. **(d)** Related to section 4.4, a part of our results that is described later in this chapter. We compare $\langle \bar{D} \rangle_t$: the time-averaged index of behavioral prediction through deviation from prototypical 1^{st} reward spike train, between $q_{opt} = 10/s$ and several lower temporal sensitivities. The average was taken over analysis windows ending between 0.1s and 1s with steps of $0.1s$. The data shown is the difference between $\langle \bar{D}(q_{opt}) \rangle_t$ and $\langle \bar{D}(q < q_{opt}) \rangle_t$. Note that for monkey M, $q = 0s^{-1}$, $q = 0.5s^{-1}$ and $q = 1s^{-1}$ lead to significantly smaller $\langle \bar{D} \rangle_t$ compared to q_{opt}, while a statistical equivalence was seen in monkey P (signed-rank test). The neurons used are the same as for Figure 4.17.

Finally, we also found similar results when decoding 1^{st} reward vs. errors (Figure 4.6 ; Figure 4.5 (c)). It is noteworthy that the improvement of decoding through temporal sensivity was also present for neurons with significant discrimination for both errors vs. 1^{st} reward, and 1^{st} reward vs. repetition (Figure 4.6 (d,e,f)). This suggests a temporal decoding advantage for a signal related to the specification of a precise behavioral strategy (exploration, switch or repetition), rather than related to the presence of reward per se.

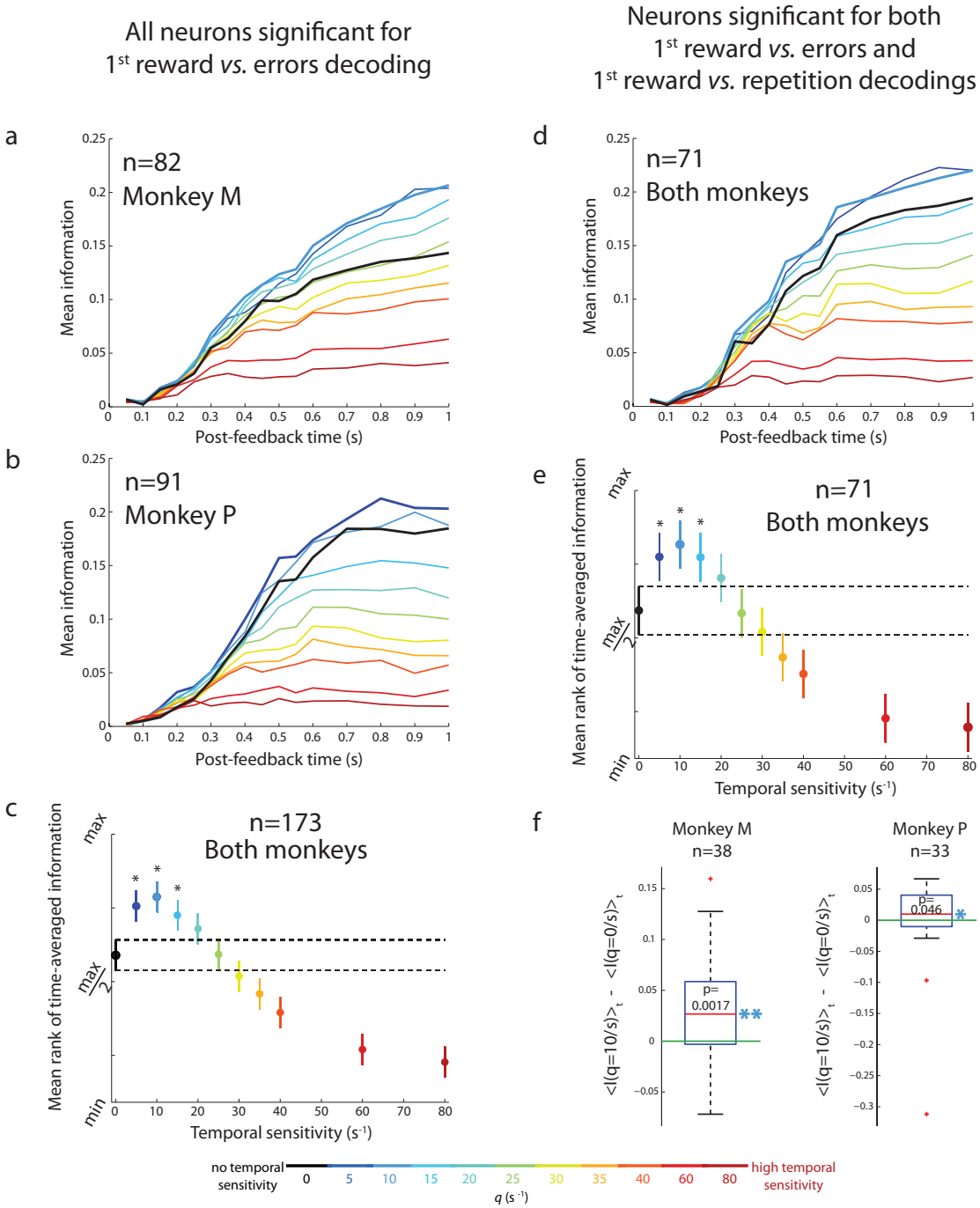


Figure 4.6: *Decoding the identity of the adapted behavioral strategy (exploration or switch).* The data suggest an advantage of spike-timing sensitivity for decoding the identity of the adapted behavioral strategy (exploration or switch). **(a,b,c)** Single unit decoding between errors and 1^{st} reward spike trains, for all neurons with significant errors *vs.* 1^{st} reward classification. **(a,b)** Time course of the mean information for different temporal sensitivities as indicated in the colorbar, for monkey M **(a)** and monkey P **(b)**. **(c)** Mean rank (±95% confidence interval) of post hoc comparisons (using Tukey's honestly significant criterion correction for multiple comparison) of a Friedman test comparing the time-averaged information $\langle I \rangle_t$. The average was taken over analysis windows ending between 0.1s and 1s with steps of 0.1 s. Data from both monkeys were pooled. **(d,e,f)** Decoding performance for errors *vs.* 1^{st} reward classification, restricted to neurons that were significant for both errors *vs.* 1^{st} reward classification and 1^{st} reward *vs.* repetition classification. The discharge of these neurons cannot therefore be merely related to the reward quantity received by the monkey, instead they appear correlated with the nature of the adapted behavioral strategy. **(d)** Time-course of mean information for different temporal sensitivities as indicated in the colorbar (data from both monkeys pooled). **(e)** Mean rank (±95% confidence interval) of post hoc comparisons (using Tukey's honestly significant criterion correction for multiple comparison) of a Friedman ANOVA comparing the time-averaged information $\langle I \rangle_t$. Note that the differences in the rankings of q-values between the mean information and the Friedman graphs are due to the fact that the mean is more sensitive to outliers with large values, while the Friedman rank is determined by the consistency (over neurons) of the within-neuron rankings of $\langle I \rangle_t$ between different q-values. These outliers are for instance visible in monkey P in **(f)**. Some of these outliers might be due to noise (e.g. the lower outlier in monkey P in **(f)** had the smallest number of trials, and less trials were available in monkey P, see Table 3.1). **(f)** Boxplots showing the distribution of the difference $\langle I(q_{opt} = 10/s) \rangle_t - \langle I(q = 0/s) \rangle_t$ for the two monkeys separately. A signed rank test was significant in both monkeys individually.

Hence, decoding of both the appropriate behavioral strategy and of the degree of necessity to update the behavior (i.e., the level of cognitive control required) could benefit from temporal sensitivity.

The curve of the amount of information vs. q was bell-shaped (Figure 4.2 (a,b), Figure 4.5). This suggests an optimal range of temporal sensitivity for decoding. If q increases further, the decoder emphasizes too much small uninformative spike time fluctuations relative to the appropriate timescale(s) of spike-timing reliability, thereby deteriorating the decoding. We were interested in comparing the range of interspike intervals occuring in the data (these intervals being computed within [0.001, 1] s post-feedback separately for all trials), and the range of interspike intervals at which temporal coincidences could occur during optimal decoding. Among significant neurons, the interquartile ranges of the median interspike interval were 54-143 and 49-110 ms for 1^{st} reward and error discrimination, respectively. Consequently, several spikes often occurred within the range of spike timing reproducibility accounted for when decoding with $q_{opt} \approx 10s^{-1}$ (i.e. a range of 200 ms, Figure 3.1 (a right)). We stress that this temporal decoder was therefore more spike-timing sensitive than a mere 200 ms binning procedure, because for $q = 10s^{-1}$, the whole range of interspike intervals between 0 and 200 ms corresponds to different values of dissimilarity. This range of interspike intervals can be interpreted as the range of presynaptic spike time jitters at which coincidences happen, i.e. leading to an effective summation of EPSPs decaying at a time scale $\tau \approx 100ms$ (Figure 3.1 (a)).

4.1.2 Temporal coding supplements, rather than competes with, spike count coding

We investigated the relation between the firing rate properties of the neurons and temporal coding. The absolute value of the difference in mean spike count between task epochs (see the definition in Table 3.3) correlated positively with the maximum time-averaged information (Spearman correlation coefficient: $c_{1^{st}\,reward} = 0.57$, $c_{errors} = 0.71$, $p < 0.001$ for all). However, large spike-count differences in highly informative neurons did not imply the absence of information related to spike timing. Indeed, among the group of neurons selected for being highly informative (through the separation of $max_q\left(\langle I\left(q\right)\rangle_t\right)$ in two clusters using a k-means algorithm), we observed an improvement of

decoding with $q \approx q_{opt}$ compared to $q = 0$ (see Figure 4.7). Also, the normalized difference in mean spike count and the gain of information related to timing sensitivity (see Table 3.3 for the definition) were negatively correlated ($c_{1^{st}\,reward} = -0.52$, $c_{errors} = -0.6$, $p < 0.001$ for all). Therefore, temporal sensitivity could uncover a relatively high amount of information in neurons with small differences in spike rate between task epochs (such as the neuron on the left of Figure 4.1 (b)).

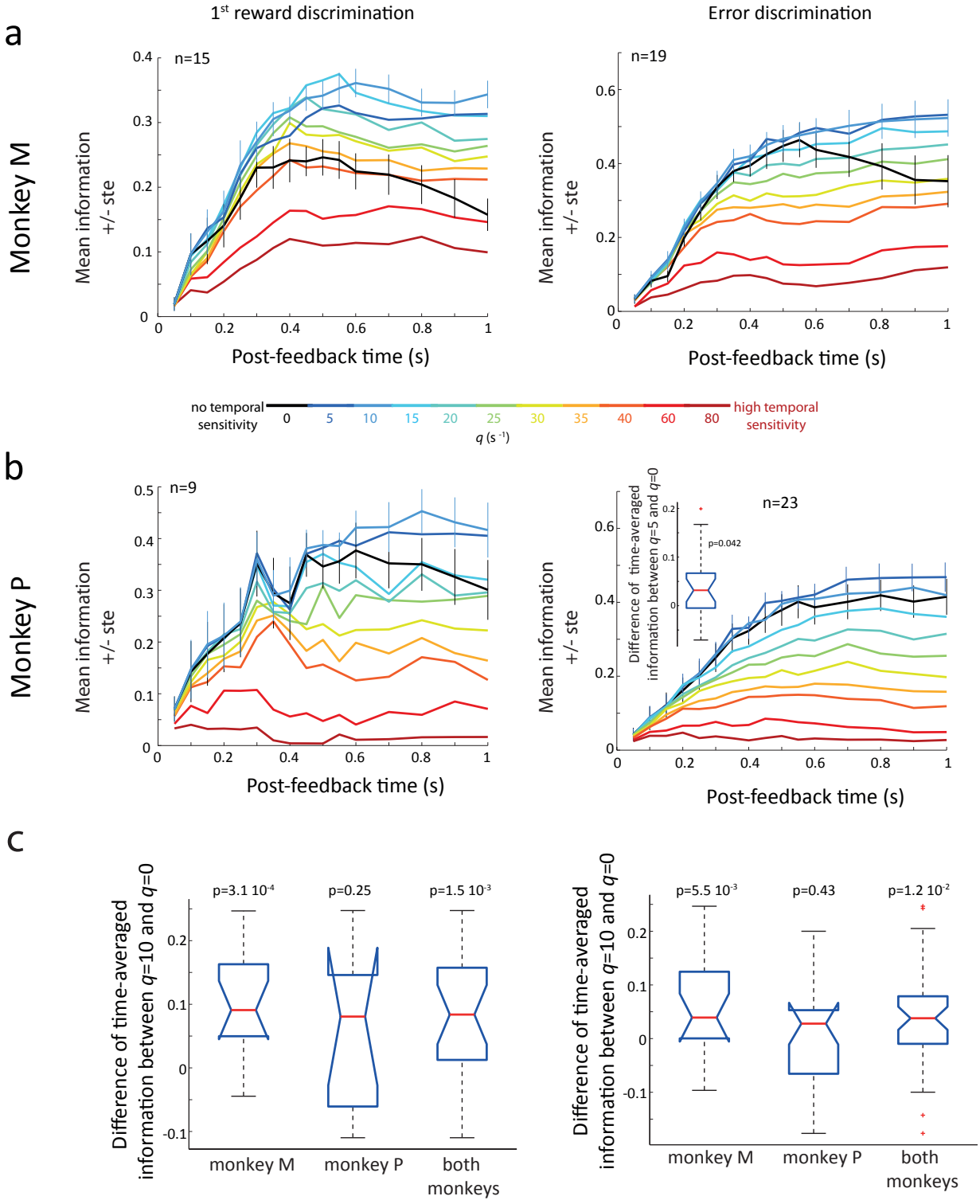


Figure 4.7: *Advantage of spike-timing-sensitive decoding over spike-count decoding for very informative neurons.* Spike-timing-sensitive decoding was also beneficial for very informative single neurons. We computed the maximum time-averaged information I_{max} for significant units (over q). Then, we used a k-means algorithm (with two groups) to separate populations with high *vs.* low I_{max}. Results in this figure are for the high I_{max} neurons. **(a,b)** show the time course of the mean information (over neurons) for 1^{st} reward (left) and errors (right) discrimination, as a function of timing sensitivity q, separately for the two monkeys. The inset in **(b, right)** shows the difference of time-averaged information $\langle I \rangle_t$ between $q = 5$ s^{-1} (found optimal for monkey P over all significant units, for errors discrimination) and $q = 0$ s^{-1}. The p-value of a signed-rank test is indicated. **(c)** boxplots of the corresponding distributions of difference in $\langle I \rangle_t$ between $q = 10$ and $q = 0$ s^{-1}. P-values of signed rank tests are indicated. Note that the notches indicate a confidence interval on the median, which may extend further away than the 25^{th} or 75^{th} quantiles, resulting in an inversion of the boxplot.

We also wondered whether the spiking activity was more reliable during some task-epochs. In order to investigate this, we took advantage of the fact that the Victor and Purpura metrics scales with the number of spikes (see section 3.2 and subsection 3.5.1). Hence, for a given neuron, the q_{opt} computed with this metric is expected to mostly reflect the spike-timing reliability of the task-epoch with more spikes, whose spike trains are harder to classify. Indeed, within spike trains of this task-epoch, a small dissimilarity d can only be reached (and therefore correct classification can only happen) if the decoder detects a very small dissimilarity per spike and therefore a sufficiently small summed dissimilarity over all spikes. Hence, we compared groups of neurons firing preferentially in different task-epochs, and found that the q_{opt} values were higher for neurons discharging more during the task-epochs requiring behavioral adaptation (see Figure 4.8 (b, left)). This difference was likely to reflect an increased reliability of spike timing during the behavioral adaptation epochs, rather than a decrease in spike count reliability, as the timing-insensitive information values ($q = 0$) were statistically indistinguishable between the groups of neurons with different firing preference (see Figure 4.8 (b, right)). For neurons firing more during repetition, optimal temporal sensitivities were distributed around $q = 5s^{-1}$. In contrast, for neurons firing more during behavioral adaptation, which were the majority, the median optimal sensitivity was $10s^{-1}$ and $7.5s^{-1}$ for 1^{st} reward and error discrimination, respectively (with a significant improvement compared to $q = 5s^{-1}$ for 1^{st} reward, see Figure 4.8 (c)). These results may reflect a higher temporal reliability of spiking during behavioral adaptation. Alternatively, our observations could also be compatible with a less reliable time reference for neural activity during repetition epochs. Indeed, the feedback could be anticipated during repetition, which may lead to a trial-specific advance of neuronal activity compared to the actual reward time.

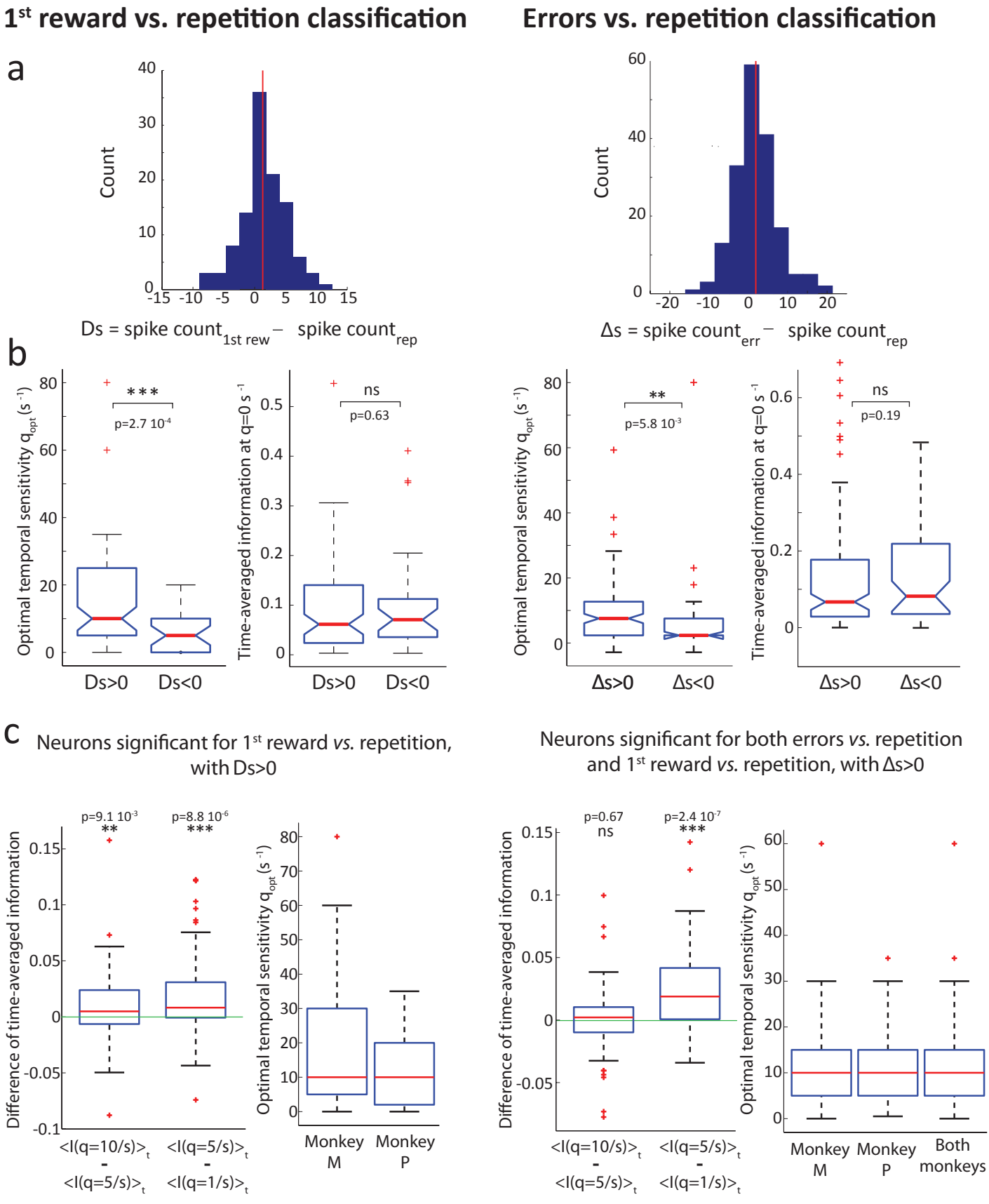

1st reward vs. repetition classification
Errors vs. repetition classification
a
Count
Ds = spike count1st rew − spike countrep
Δs = spike counterr − spike countrep
b
Optimal temporal sensitivity qopt (s-1)
Time-averaged information at q=0 s-1
***
p=2.7 10-4
ns
p=0.63
**
p=5.8 10-3
ns
p=0.19
Ds>0
Ds<0
Δs>0
Δs<0
c
Neurons significant for 1st reward vs. repetition, with Ds>0
Neurons significant for both errors vs. repetition and 1st reward vs. repetition, with Δs>0
Difference of time-averaged information
p=9.1 10-3
**
p=8.8 10-6
***
p=0.67
ns
p=2.4 10-7
***
<I(q=10/s)>t - <I(q=5/s)>t
<I(q=5/s)>t - <I(q=1/s)>t
Monkey M
Monkey P
Both monkeys

Figure 4.8 ***(previous page)*:** *The optimal decoding temporal sensitivity appeared higher for neurons firing more during behavioral adaptation. Left*: 1^{st} reward *vs.* repetition discrimination; *right*: errors *vs.* repetition discrimination. **(a)** Difference of mean spike count in a $[0.001, 1]$ s post-feedback window between behavioral adaptation and repetition epochs, across neurons with significant discrimination between task-epochs. The vertical red line marks the median of the distribution. **(b, Left)** Boxplots of the distributions of q_{opt} values for cells discharging preferentially during behavioral adaptation *vs.* repetition (the notches indicate an approximate confidence interval on the median, which may extend beyond the quartiles). P-values of ranked sum tests comparing medians are shown. q_{opts} values were larger for neurons firing more during behavioral adaptation. **(b, Right)** Boxplots of the distributions of time-averaged information $\langle I \rangle_t$ for $q = 0$ s^{-1}. P-values of ranked sum tests are shown. The absence of significant difference suggests that the difference in q_{opt} (left) reflects a difference in spike-timing reliability rather than a difference in spike-count reliability between the groups. **(c)** Detailed analysis about the optimal temporal sensitivity q_{opt} for decoding cognitive-control signals during feedbacks of the task which should trigger behavioral adaptation (1^{st} reward or errors). We focus on neurons discharging more during 1^{st} reward for 1^{st} reward *vs.* repetition discrimination and on neurons discharging more during errors for errors *vs.* repetition discrimination. In addition, for errors *vs.* repetition discrimination, we focus on cognitive control coding and exclude putative 'physical reward' coding by only selecting neurons that were significant for both errors *vs.* repetition and 1^{st} reward *vs.* repetition (n=32 from monkey M, n=27 from monkey P). **(c, Left)** Difference of time-averaged information $\langle I \rangle_t$ between $q = 10/s$ and $q = 5/s$, and between $q = 5/s$ and $q = 1/s$; the p-value of a signed-rank test for the distribution of the difference values around 0 is shown. **(c, Right)** Distribution of optimal temporal sensitivities (here, including data at $q = 0.5/s$ and $q = 1/s$) showing that for both discriminations and for both monkeys independently, the median q_{opt} was 10/s. For errors *vs.* repetition, the pooled distribution over monkeys (concerning a subset of neurons compared to (b)) is also shown.

4.1.3 Sensorimotor differences between task epochs are not likely to determine the advantage of temporal decoding

Sensory or motor differences between task epochs were unlikely to determine the advantage of temporal decoding. In fact, external events (e.g., feedback, stimuli) were identical during 1^{st} reward and repetition epochs. As we will now explain, the motor influence on neural activity was also unlikely to cause temporal decoding advantage through a different timing of eye-movements in the two task epochs (detailed methods are in section 3.4, analysis possible in monkey M). This is in agreement with the current views of dACC function [Shenhav et al. (2013)].

A control for motor correlates was still necessary in our task, as the monkeys were not forced to maintain fixation after target touch, and they often broke fixation before one second post-reward. Further, there were consistent differences in the timing of eye movements between task-epochs (see Figure 4.9 (a-b) for monkey M).

If the temporal structure of dACC activity were merely motor related, all eye

movements timed differently between task epochs would favor temporal coding. To test this hypothesis, we removed all trials with an eye movement occurring before 0.65-0.85 s post-feedback, and kept the remaining trials. More precisely, we removed the trials either if an eye-movement was detected before the end of the neuronal analysis window (hence removing putative motor-feedback activity), or if an eye movement was detected before a time corresponding to the end of the analysis window plus 300 ms (hence removing also putative premotor activity). This manipulation did not decrease the advantage of temporal decoding of 1^{st} reward vs. repetition (Figure 4.9 (d-e).: removing putative motor-feedback activity; Figure 4.9 (f-g): removing also putative premotor activity). Following target fixation, late 1^{st} eye movements ($\approx$ 850 ms after 1^{st} reward delivery) also predicted that monkeys would be quicker to respond in the following trial (see Figure 4.9 (c)). Therefore, dACC neural activity occurring either before or during these late eye movements may not reflect motor planning but rather cognitive correlates (e.g., attentional modulation). These trials with late 1^{st} eye movements indeed appeared to contribute to the temporal advantage for decoding (see analysis windows $\geq$650 ms in Figure 4.9 (d-j)). In other words, while there was an interaction between eye movement and temporal coding in dACC activity, the relation was unlikely to reflect the presence of a pattern of activity in dACC triggering eye movements; rather, dACC activity and eye movements appeared to be both modulated by cognitive control.

In consequence, we note that the different temporal patterns between 1^{st} reward and repetition task epochs probably originated from different internally generated neuronal dynamics.

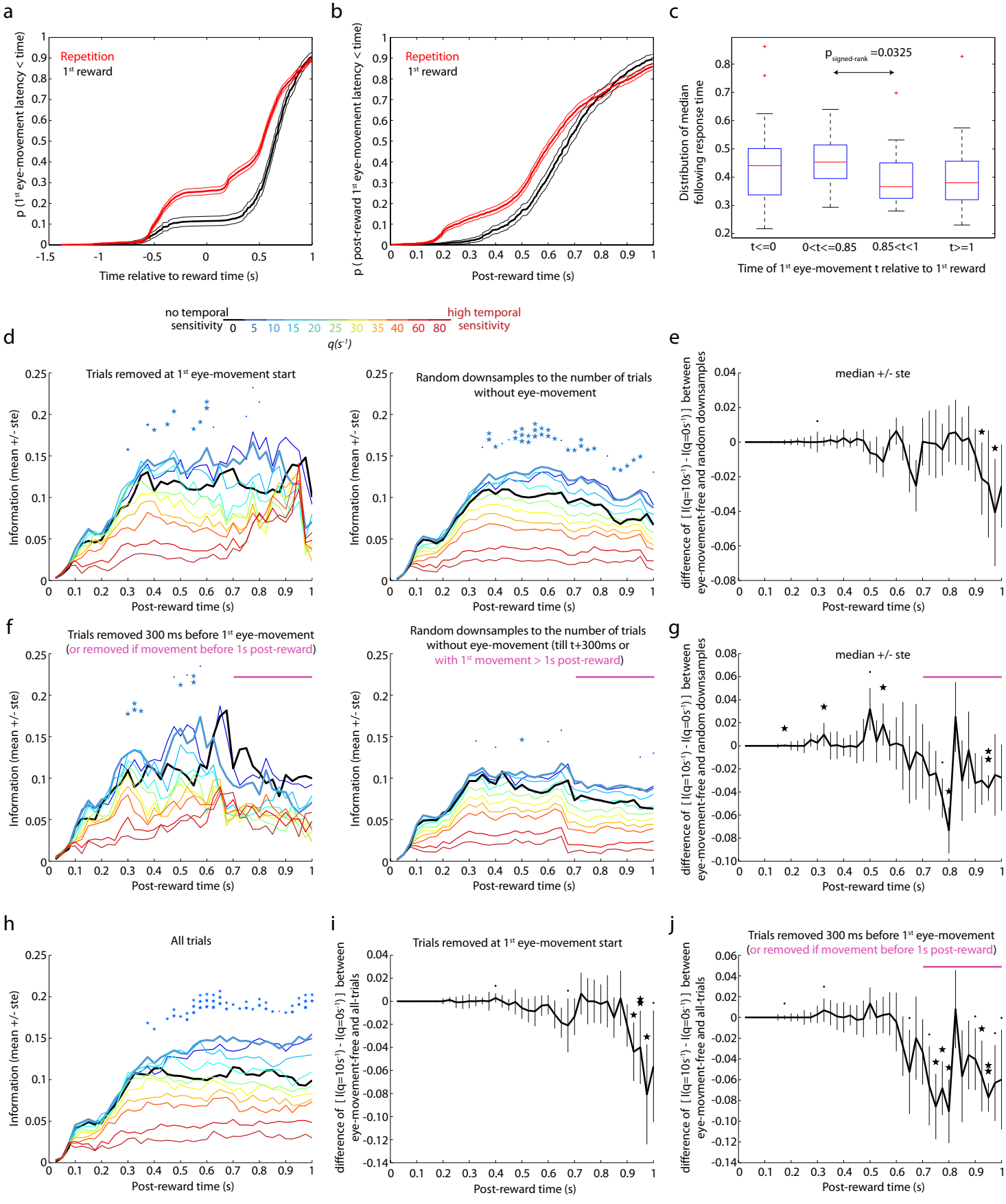

a
Repetition
1st reward
p (1st eye-movement latency < time)
Time relative to reward time (s)
b
Repetition
1st reward
p (post-reward 1st eye-movement latency < time)
Post-reward time (s)
c
p signed-rank =0.0325
Distribution of median following response time
t<=0
0<t<=0.85
0.85<t<1
t>=1
Time of 1st eye-movement t relative to 1st reward
no temporal sensitivity
high temporal sensitivity
0 5 10 15 20 25 30 35 40 60 80
q(s-1)
d
Trials removed at 1st eye-movement start
Information (mean +/- ste)
Post-reward time (s)
Random downsamples to the number of trials without eye-movement
e
median +/- ste
difference of [I(q=10s-1) - I(q=0s-1)] between eye-movement-free and random downsamples
f
Trials removed 300 ms before 1st eye-movement
(or removed if movement before 1s post-reward)
Random downsamples to the number of trials without eye-movement (till t+300ms or with 1st movement > 1s post-reward)
g
h
All trials
i
Trials removed at 1st eye-movement start
difference of [I(q=10s-1) - I(q=0s-1)] between eye-movement-free and all-trials
j
Trials removed 300 ms before 1st eye-movement
(or removed if movement before 1s post-reward)

Figure 4.9 ***(previous page)*:** *Decoding trials without eye-movements (monkey M).* In this figure, small dot indicates $p < 0.1$, one star: $p < 0.05$, two stars: $p < 0.01$ for signed-rank tests. Error bars are standard error of the mean/median. **(a,b,c)** Behavior for 28 sessions (during which we recorded significant 1^{st} reward *vs.* repetition). **(d,e,f,g,h,i,j)** Decoding; (d,e,i) are related to the putative influence of motor activity and (f,g,j) to the putative influence of premotor activity. 38 neurons were available for analysis windows $\leq$425 ms at least; for longer windows some neurons were excluded because no trials free of saccades were available. **(a)** Cumulative distribution function of 1^{st} eye-movement latency following the fixation period. 95% confidence interval use Greenwood's formula. **(b)** As **(a)** but restricted to post-reward 1^{st} eye-movement latency. **(c)** Distributions (over different behavioral sessions) of median response times at the trial following 1^{st} reward depending on the 1^{st} eye-movement latency after the fixation period leading to 1^{st} reward. **(d)** *L*eft: Neuron-averaged information, including only trials with no eye-movements detected before the end of each analysis window. P-values compare between $q = 10\ s^{-1}$ and $q = 0\ s^{-1}$. *R*ight: Neuron-averaged information for random downsamples (from all data) to the trial numbers of (d Left). The downsampling aims at excluding a possible effect of trial number when comparing data without (left) and with (right) saccades. For each neuron, the mean information among 1000 downsamples was taken (taking the median gives similar results). P-values compare between $q = 10\ s^{-1}$ and $q = 0\ s^{-1}$. Note that the smoother aspect of the curves compared to the left graph likely results from the presence of an additional downsampling-averaging in the right graph. Note also that, here, until plateau is reached ($\approx$ 600 ms post-feedback), there were no robust differences in spike-count based information between eye-movement free and resampled data (signed-rank test on time-averaged information between 0 and 600ms, or 300 and 600 ms, all $p_s > 0.16$). **(e)** Median difference of information increase thanks to temporal structure: $[I(q = 10\ s^{-1}) - I(q = 0\ s^{-1})]$, between eye-movement-removed (d Left) and randomly downsampled data (d Right). Negative values indicate a smaller timing-sensitivity-related improvement in decoding for eye-movement-free data. **(f)** Conventions as in (d). *L*eft: Only trials for which 1^{st} eye-movement occured later than ((analysis window end)+300ms) were included. Due to limitations in trial number, for analyses windows longer than 700 ms, all trials with 1^{st} eye-movement latency $\geq$1 s post-reward were included. *R*ight: data randomly downsampled to the trial number of (f Left). **(g)** Difference of information increase thanks to temporal structure: $[I(q = 10\ s^{-1}) - I(q = 0\ s^{-1})]$, between eye-movement-removed (f Left) and randomly downsampled data (f Right). **(h)** Neuron-averaged information among all 38 available neurons, all trials included. **(i)** Difference of information increase thanks to temporal structure: $[I(q = 10\ s^{-1}) - I(q = 0\ s^{-1})]$, between eye-movement-removed as in (d Left), and total data. **(j)** Difference of information increase thanks to temporal structure: $[I(q = 10\ s^{-1}) - I(q = 0\ s^{-1})]$, between eye-movement-removed as in (f Left), and total data.

4.2 Temporal decoding of 1^{st} reward vs. repetition spiking does not only rely on differences in time-varying firing rate between task epochs

We investigated the nature of dACC firing statistics determining the advantage of temporal decoding. Spike-timing reliability might mainly reflect differences in the temporal variations of firing rates between task epochs. Alternatively, beyond this time-dependent firing rate, temporal correlations between spikes within one

trial may impact the spike time reproducibility. Indeed, cellular processes (such as spike-triggered hyperpolarizing currents or short-term plasticity) may lead to a dependence of future spiking probability on past spike times [Arsiero et al. (2007); Mongillo et al. (2008); Schwalger and Lindner (2013)], in particular if the synaptic current received by the neuron is not very variable. Similarly, recurrent neural network dynamics – within dACC or upstream – may create correlations in spike times [Brunel (2000); Ostojic (2014)]. Here, we tested whether or not, beyond their existence, spike-timing correlations sizably and consistently (over neurons) impacted information transmission.

4.2.1 Assuming a time-dependent firing rate implies a spike count variability incompatible with the data

We randomly shuffled spike times within each task epoch while preserving the peri-event time histograms (PETH) for each neuron (see Figure 4.10 (a)). This transformation preserved the time-dependent firing rate, while destroying temporal correlations. It also created an approximate "Poisson" spike count variability, i.e. a variability that was purely determined by random samples from a unique time-dependent firing probability (see subsection 3.2.2).

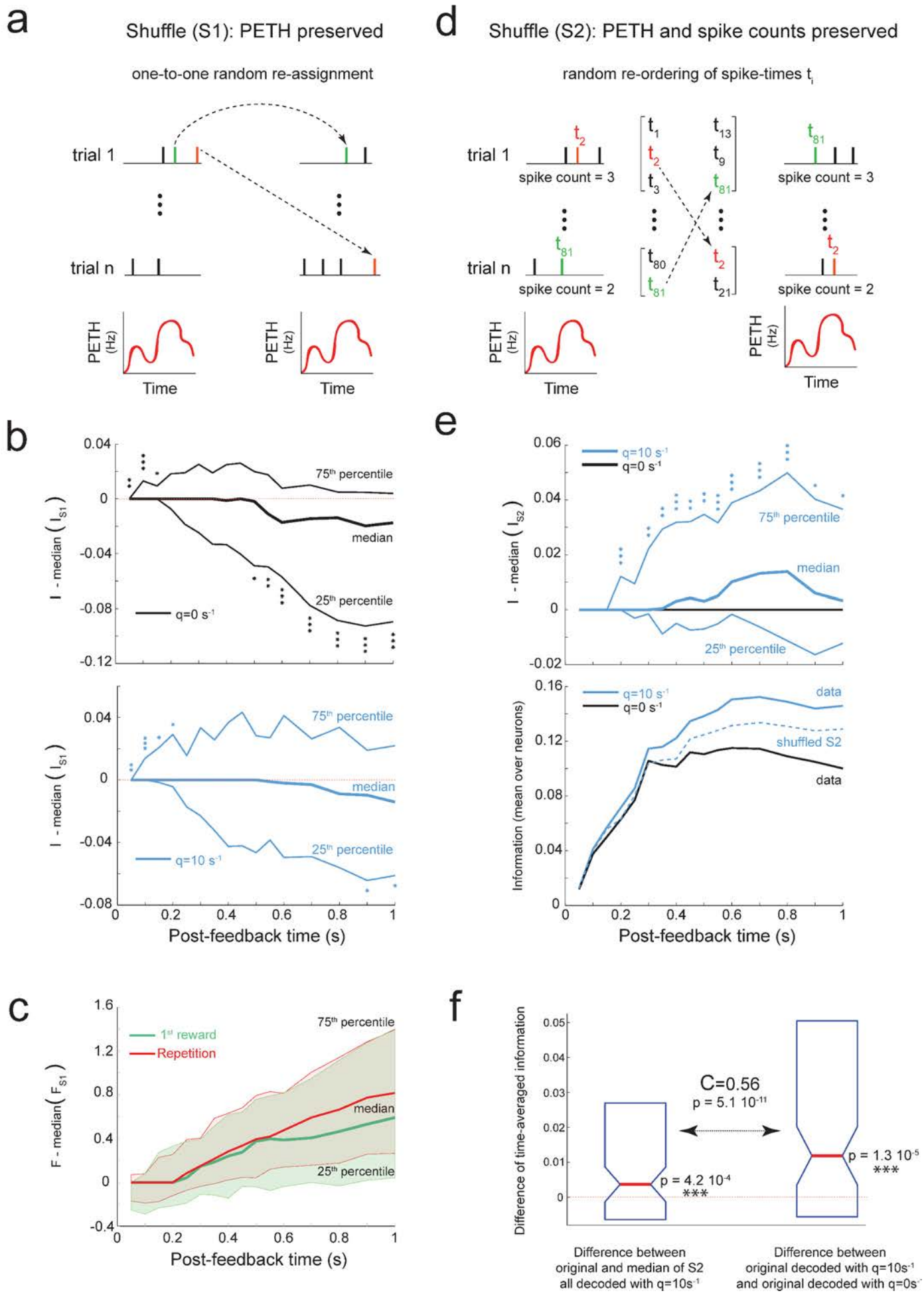

a
Shuffle (S1): PETH preserved
one-to-one random re-assignment
trial 1
trial n
PETH (Hz)
Time
d
Shuffle (S2): PETH and spike counts preserved
random re-ordering of spike-times t_i
spike count = 3
spike count = 2
b
I - median (I_{S1})
75th percentile
median
25th percentile
q=0 s-1
q=10 s-1
Post-feedback time (s)
e
I - median (I_{S2})
Information (mean over neurons)
data
shuffled S2
c
F - median (F_{S1})
1st reward
Repetition
f
Difference of time-averaged information
C=0.56
p = 5.1 10-11
p = 4.2 10-4
***
p = 1.3 10-5
Difference between original and median of S2 all decoded with q=10s-1
Difference between original decoded with q=10s-1 and original decoded with q=0s-1

Figure 4.10 ***(previous page)*:** *Temporal decoding does not only rely on differences in time-varying firing rate.* This analysis was restricted to neurons significantly discriminating between 1^{st} reward and repetition task epochs. **(a)** Shuffling of spikes between trials while preserving PETHs (i.e. time-dependent firing rates). This procedure was repeated 1000 times within each task epoch and independently for each neuron. If information transmission in the data relies on PETHs, spike shuffling should not impact decoding. **(b)** Distribution of the difference between information I in original data and the median information in shuffled data (as in **a**). Stars indicate the significance of a signed-rank test: 1 to 3 stars, $p \leq 0.05$, $p \leq 0.1$, $p \leq 0.001$, respectively. Top: spike-count decoding ($q = 0s^{-1}$). Bottom: optimal temporal decoding at $q_{opt} \approx 10s^{-1}$ (this q value maximized information averaged over neurons). **(c)** Difference of Fano factor estimate (F, defined in Table 3.3) between original data and the median of 1000 shuffled data sets for 1^{st} reward (green) and repetition (red). **(d)** Shuffling of spikes between trials while preserving both PETHs and spike count variability. The shuffling is done 1000 times, independently for all analysis windows, task-epochs and neurons. All spikes emitted during different trials of a task epoch are grouped and their order shuffled. Each pseudo-trial (right) is created by taking from the shuffled spike pool (middle) the same number of spikes as in the corresponding original trial (left). If information transmission in the data were shaped by a PETH time-course whose total integral could change across trials, spike shuffling would not impact decoding. **(e)** Top: Distribution of the difference between information in original data and the median information in shuffled data (as in **d**). Note that for $q = 0s^{-1}$ the curves of median, 25^{th} and 75^{th} percentiles are overlapped. Bottom: mean information in the original data decoded with $q_{opt} \approx 10s^{-1}$ and with q=0 s^{-1}, and in spike trains shuffled (as in d) decoded using $q_{opt} \approx 10s^{-1}$. (f) Left boxplot: difference between time-averaged information $\langle I \rangle_t$ in original data and the median of $\langle I \rangle_t$ in shuffled data (as in d) at $q_{opt} \approx 10s^{-1}$, with signed-rank p-value. Right boxplot: for comparison, the difference in time-averaged information between q_{opt} and $q = 0s^{-1}$ in original data. Box plots show 25^{th}, 50^{th} and 75^{th} percentiles. The two quantities (left and right boxplots) were correlated (with coefficient C).

If information transmission were shaped by time-dependent firing rates, original and spike-shuffled data should convey similar information. In contrast, we found that both spike-count and timing-sensitive decoding at $q_{opt} \approx 10s^{-1}$ were more reliable for short analysis windows, and less reliable for long analysis windows, in the original compared to shuffled data (Figure 4.10 (b)). These results were consistent and robust in both monkeys (Figure 4.11 (a)).

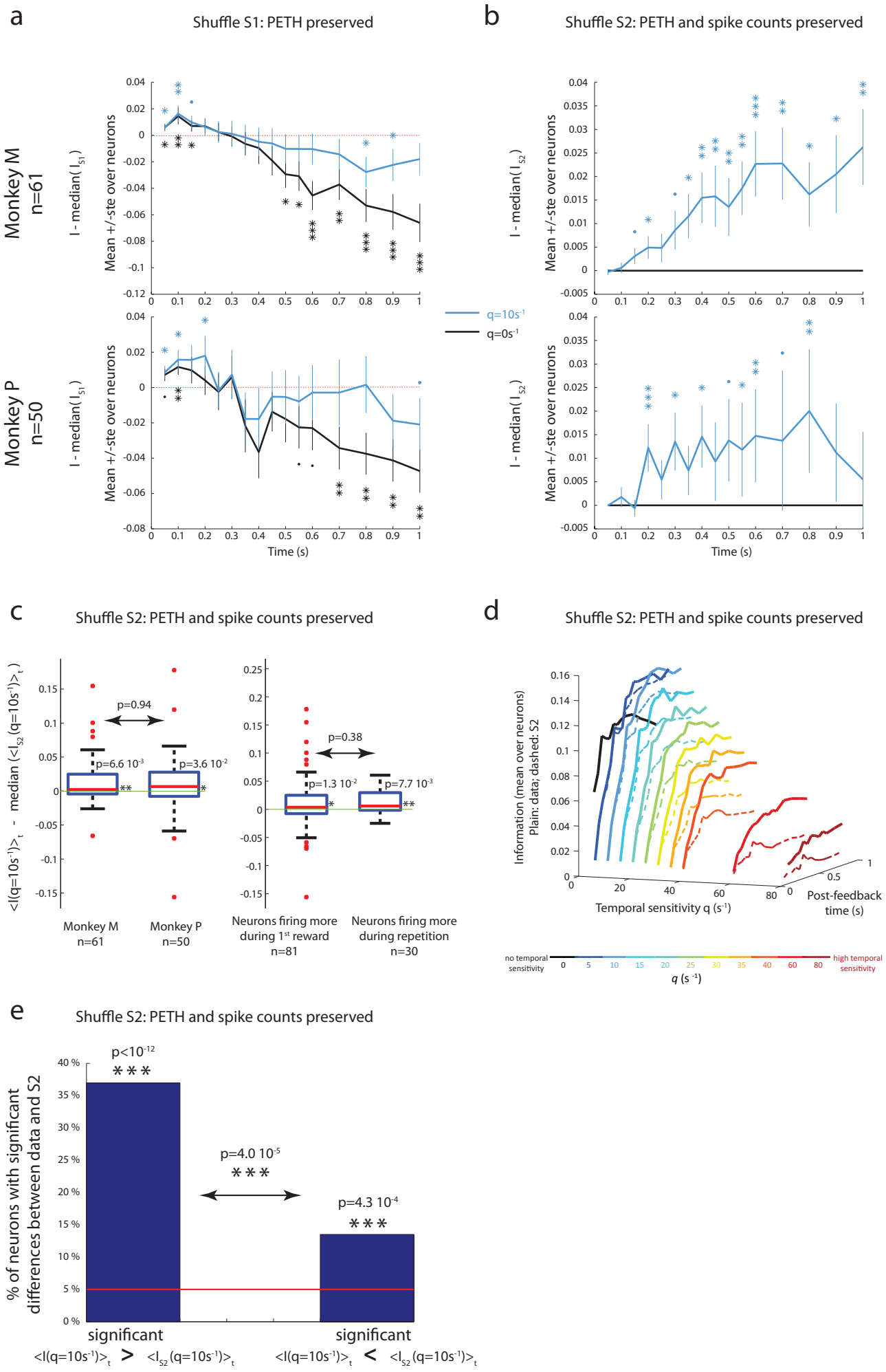

a
Shuffle S1: PETH preserved
Monkey M n=61
Monkey P n=50
I - median(I_S1)
Mean +/-ste over neurons
Time (s)
q=10s-1
q=0s-1
b
Shuffle S2: PETH and spike counts preserved
I - median(I_S2)
Mean +/-ste over neurons
Time (s)
c
Shuffle S2: PETH and spike counts preserved
<I(q=10s-1)>t - median (<I_S2(q=10s-1)>t)
p=0.94
p=6.6 10-3
p=3.6 10-2
p=0.38
p=1.3 10-2
p=7.7 10-3
Monkey M n=61
Monkey P n=50
Neurons firing more during 1st reward n=81
Neurons firing more during repetition n=30
d
Shuffle S2: PETH and spike counts preserved
Information (mean over neurons) Plain: data; dashed: S2
Temporal sensitivity q (s-1)
Post-feedback time (s)
no temporal sensitivity
high temporal sensitivity
q (s -1)
e
Shuffle S2: PETH and spike counts preserved
% of neurons with significant differences between data and S2
p<10-12
***
p=4.0 10-5
***
p=4.3 10-4
***
significant
significant
<I(q=10s-1)>t > <I_S2(q=10s-1)>t
<I(q=10s-1)>t < <I_S2(q=10s-1)>t

Figure 4.11 ***(previous page)*****:** *Robustness of the link between spiking statistics and information transmission.* **(a)** The changes in information induced by performing shuffle 1 (preserving the time-dependent rate,see Figure 4.10 (a)) were consistent over monkeys and were following the time-course of the fano factors (see Figure 4.10 (c)). For long analyses windows, original data were less reliable than their spike-shuffled counterparts, while this effect was inverted for short analysis windows. The curves are the mean +/- standard error (ste, among all significant neurons for 1^{st} reward vs. repetition classification) of the difference between the information in the original data and the median information of the corresponding shuffled data sets. We show $q = 0$ (spike-count decoding, black) and $q = 10s^{-1} \approx q_{opt}$ (blue). **(b)** Same conventions as in **(a)**. The change in information induced by shuffling spikes according to shuffle 2 (preserving both time-dependent rate and spike count variability, see Figure 4.10 (d)) were consistent over monkeys. Original data had higher information than their spike-shuffled counterparts. **(c)** The distribution of difference of time-averaged information ($< I(q = 10s^{-1} \approx q_{opt}) >_t$) between original data and the median for the corresponding data sets created by shuffle 2 was significantly positively biased for both monkeys (left) and for both the neurons firing more during 1^{st} reward and the neurons firing more during repetition (signed-rank tests, all $p_s < 0.036$). Note that q_{opt} is unambiguously $10s^{-1}$ for neurons firing more during 1^{st} reward (for these neurons $q = 5s^{-1}$ and $q = 15s^{-1}$ perform very similarly for original data decoding, see also Figure 4.8 (c)). The distributions were not different between monkeys or between firing preference (ranked-sum tests, all $p_s > 0.38$). **(d)** We show the means (over neurons) of (i) the information in original data, and of (ii) the median information of the corresponding shuffled data sets. For all q values, we observed higher information for the original data as compared to their shuffle 2 counterparts. The size of the effect increased for higher q values. **(e)** The proportion of neurons for which shuffle 2 led to a significant decrease in $< I(q = 10s^{-1}) >_t$ (more than 95 % of shuffled data sets with smaller $< I(q = 10s^{-1}) >_t$ than original, left), was higher than the proportion of neurons with a significant increase (more than 95 % of shuffled data sets with larger $< I(q = 10s^{-1}) >_t$ than original, right). Proportions were compared using the Fisher's Exact Probability Test with mid-p correction ($p = 4.0\ 10^{-5}$). Both of these proportions were larger than chance (5%): binomial test, all $p_s < 10^{-3}$.

Timing-sensitive and spike-count decoders were both impacted by changes in spike count variability.

For short analysis windows, the improved reliability of spike count in the original data could be linked to spike-triggered hyperpolarizing currents which can counterbalance random deviations of neuronal excitability in single neurons [Arsiero et al. (2007); Farkhooi et al. (2011); Schwalger and Lindner (2013)]. This increased spike count reliability compared to Poisson firing is actually more likely to happen if the neurons receive an input current which fluctuates little (even in neurons for which spikes only trigger a simple reset of the voltage and an absolute refractory period, see [Litwin-Kumar and Doiron (2012)] for instance).

For long analysis windows, the spike count appeared more variable in the original data (as measured by the Fano factor; Figure 4.10 (c)), causing a smaller decoding reliability. This means that spike count variability in the original data cannot be explained by random samples taken from a single firing probability. More precisely, for post-feedback times longer than 500 ms, the spiking probability

was actually stronger in some trials than in other trials. This suggests a hidden source of spike count variability across trials which is not constant during one task epoch and which has a major influence on information transmission [Litwin-Kumar and Doiron (2012); Ostojic (2014)]. This large spike count variability may reflect the integrative properties of dACC. Indeed, beyond signaling the need for behavioral adaptation, dACC firing may also be influenced by factors such as attention [Totah et al. (2013)] and/or target identity [Procyk et al. (2000)]. Interestingly, this large spike count variability appeared to hinder more spike count decoding (Figure 4.10 (b)). Hence, this may have participated to shaping the larger difference of information between $q_{opt} \approx 10s^{-1}$ and $q = 0s^{-1}$ decoders which occurred for long analysis windows ($\geq 500ms$) compared to short windows (Figure 4.2 (a-c)).

4.2.2 Temporal correlations considerably impact information transmission

We tested whether the information transmission could be mainly shaped by the combination of the PETH time-courses and of the spike-count variability of the data. To do so, we shuffled spike times while preserving both PETHs and spike counts in all trials (Figure 4.10 (d)). Through this operation, spike-count information was conserved in the shuffled data. If temporal correlations had negligible impact on information transmission, then temporal decoding should also remain unchanged. In contrast, we found that for $q_{opt} \approx 10s^{-1}$, information decreased in the shuffled data as compared to original ones (Figure 4.10 (e-f)). These results were robust and consistent across monkeys (Figure 4.11 (b-c)). The temporal correlations of the original data increased information by about 10-15%, on average, compared to shuffled data (Figure 4.10 (e), information at plateau). In addition, the increase of information with optimal temporal sensitivity q_{opt} (compared to spike count) was significantly correlated to the information increase with temporal correlations (Figure 4.10 (f)). This further suggests that temporal correlations tended to support temporal coding. Finally, information loss after spike shuffling was larger for larger temporal sensitivities q (Figure 4.11 (d)), suggesting that spike correlations were stronger at shorter time-scales.

Altogether, our results suggest that, beyond the time-dependent firing rates, temporal correlations led to spike timing reliability that favored task-epoch discrimination in most neurons. The neurons for which correlations appeared to

make the neural activity more similar between task-epoch, and therefore to impair the classification of neuronal responses in different task epochs, were therefore the minority (Figure 4.11 (e)).

These results could reflect either the single neurons dynamics [Arsiero et al. (2007); Pozzorini et al. (2013); Park et al. (2014)], or network dynamics mediating the behavioral strategy signal, that would make future spike times dependent on spiking history [Brunel (2000); Ostojic (2014)].

4.3 Temporal patterns often differ between neurons, implying a spatiotemporal code

Multiple neurons were often simultaneously recorded (median=2). Thus, we also decoded the activity of pairs of neurons (Monkey M, n=122 pairs; Monkey P, n=271) while varying both the temporal sensitivity q and the degree of distinction between neurons k (see Figure 3.1 (b), subsection 3.2.1). For the computation of the dissimilarity measure, the parameter k represents the cost of transforming a spike from neuron 1 into a spike from neuron 2. Therefore, during classification of spike trains from pairs of units, the dissimilarity between spikes from different neurons increases with k. The parameter k permits to test whether the informative spikes are neuron specific, or if they are emitted synchronously by two neurons. In the former case, the amount of information would increase if the decoder were accounting for neural identity ($k > 0$), as compared to a decoder blind to neural identity and sensitive to interferences between neurons ($k = 0$). In the latter case, $k = 0$ could be optimal for decoding because it makes the discharge of either one of the neurons sufficient to have reliable joint spiking. Note that for this situation to occur, the discharge of informative spikes should not be strongly positively correlated between the two neurons (else, the signals emitted by the two neurons are redundant and cannot complement one another).

4.3.1 Paired decoding benefits from an optimal distinction between the spikes from the two neurons

We first tested whether decoding with optimal (q, k) values advantageously combined the activity of any two analyzed neurons (regardless of their individual coding properties). This was not trivial because of the imbalance in information between neurons (Figure 4.12 (a)). Also, in our data, when some "noise" (relative to the mean task-epoch "spike-count signal") caused one neuron to fire more, it was not in general causing the other neuron to fire less. Indeed, the spike counts emitted during a task-epoch were not negatively correlated in our data. Thus, summing the activity of two neurons would not cancel the effect of noise on spike counts. Hence, the (wide) distribution of spike-count correlations between two neurons was slightly positively biased during 1^{st} reward or repetition (signed-rank test on time-averaged correlation coefficients: $p = 1.6\ 10^{-3}$ with median 0.043 for 1^{st} reward; $p = 1.4\ 10^{-5}$ with median 0.036 for repetition). During errors, the distribution of correlation coefficients was centered on zero.

In general, a simple sum of two independent or positively correlated neural activities which are associated with largely different standard deviations is likely to decrease the signal-to-noise ratio, compared to the more reliable single activity. In contrast, the decoding relying on the multi-unit dissimilarity measure most often uncovered more information in a pair compared to the most informative neuron of the pair ("gain in the pair relative to the best single unit", as defined in Table 3.3 ; Figure 4.12 (a), signed-rank test, all $p_s < 0.001$). As expected, information gains were negatively correlated with the information imbalance between paired neurons (Figure 4.12 (a), Spearman correlation with permutation test, all $p_s < 0.001$; more pronounced for pairs with significant coding: Figure 4.13).

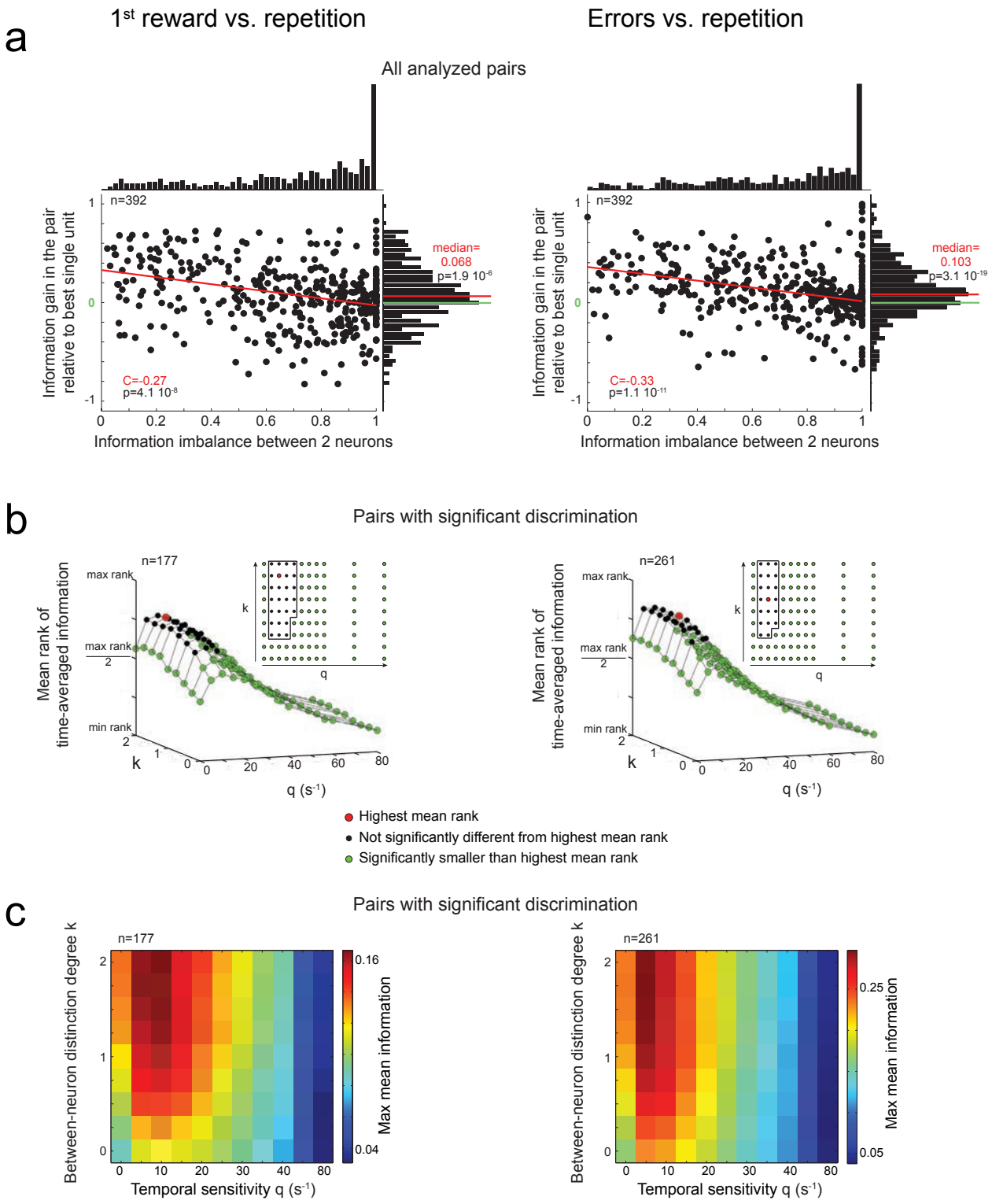


Figure 4.12: *Efficient paired decoding often required to distinguish between the activities of the two neurons. Left*: Decoding 1^{st} reward *vs.* repetition task epochs. *Right*: Decoding error *vs.* repetition task epochs. **(a)** Distribution of information gain when decoding a pair of units relative to decoding the isolated unit of the pair with the highest information, as a function of the information imbalance between the two units of the pair (defined in Table 3.3). The red line indicates a linear regression fit. The distributions of information gains were significantly biased toward positive values, as indicated by a signed-rank test (all $p_s < 10^{-5}$). **(b)** Mean rank comparison (with Friedman ANOVA) of the time-averaged information $\langle I \rangle_t$ as a function of (q,k). Data were pooled from both monkeys and were restricted to pairs with significant information. **(c)** Maximum mean (over neurons with significant information) information as a function of (q,k). Information was maximized over analysis windows ending in [0.05, 0.6] s, steps of 50 ms, and in [0.7, 1] s, steps of 100 ms.

We then investigated which (q, k) values yielded better dACC decoding. For any k value, the time-averaged information $\langle I \rangle_t$ significantly increased with temporal sensitivity up to $q_{opt} \approx 10s^{-1}$ and decreased for larger q values (Figure 4.12 (b-c)). Hence, for any value of k, spike-count decoding ($q = 0s^{-1}$) led to a significantly lower $\langle I \rangle_t$ than optimal temporal sensitive decoding ($q_{opt} \approx 10s^{-1}$). $\langle I \rangle_t$ also increased with k and plateaued at about 1. Therefore, intermediate to high levels of distinction between spikes from paired neurons often improved the decoding of behavioral adaptation signals, suggesting that some reliable spikes were neuron specific. Differences in information average (over significant pairs) across (q, k) values were consistent over time and between monkeys (Figure 4.14).

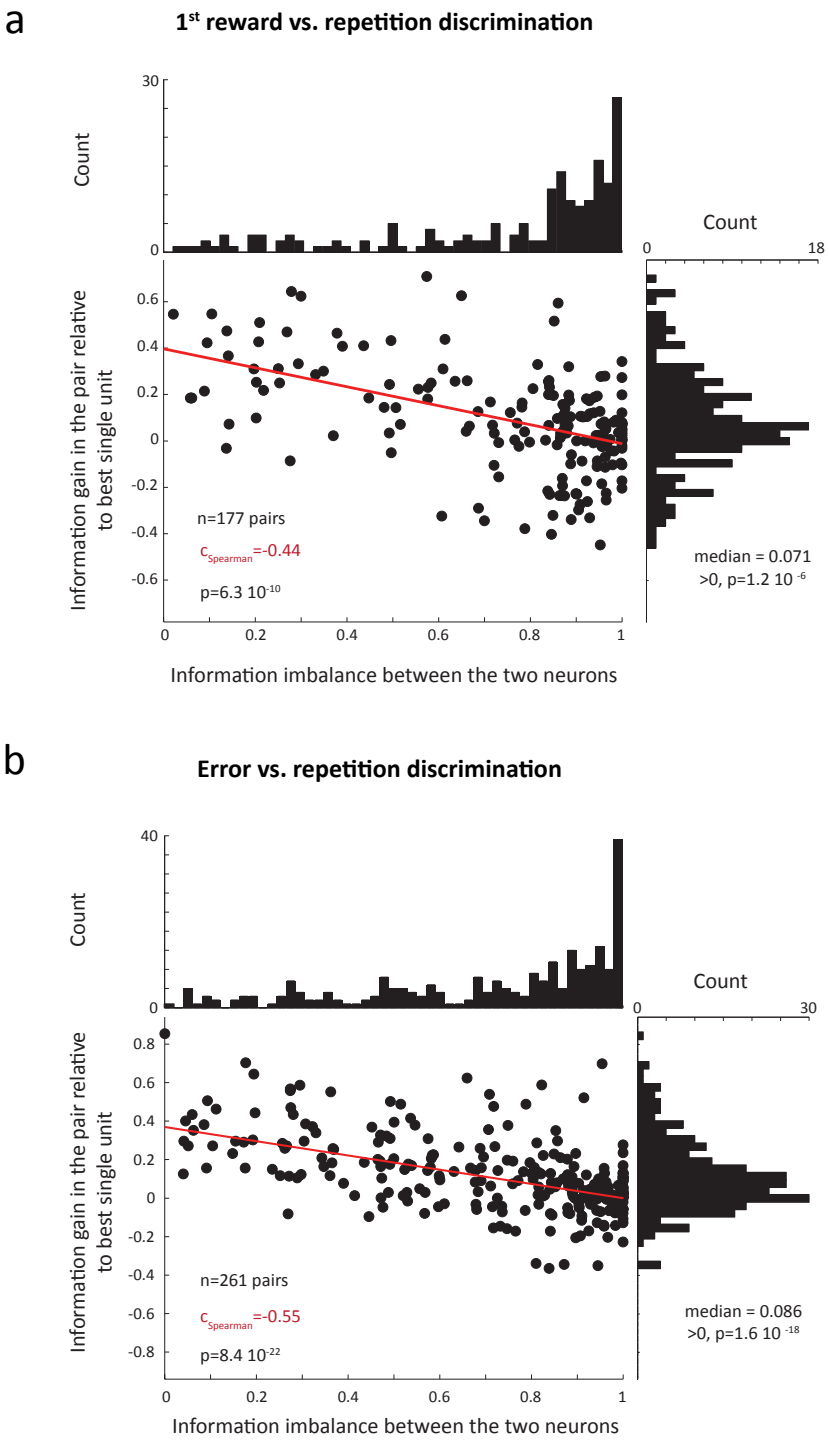


Figure 4.13: *Gains of information among pairs of neurons with significant information.* Paired spatial decoding led to increases in the amount of information despite imbalances in the discriminative power of single units. In this figure, only pairs with significant classification (permutation test) were included. **(a)** Discrimination between first reward and repetition task-epochs. The central plot shows the correlation between the information gain (obtained when decoding a neuron pair *vs.* the pair's most informative single unit, see Table 3.3) and the degree of information imbalance between the two units of a pair. The p-value testing whether the correlation differed from 0 is indicated ($p < 0.001$). The red line is a linear fit. The histograms at the top and right show the two marginal distributions. A signed-rank test was used to measure the significance of the bias towards an increase in the amount of information (i.e. positive gains, $p < 0.001$). **(b)** Same as (a) but for the discrimination between error and repetition task-epochs.

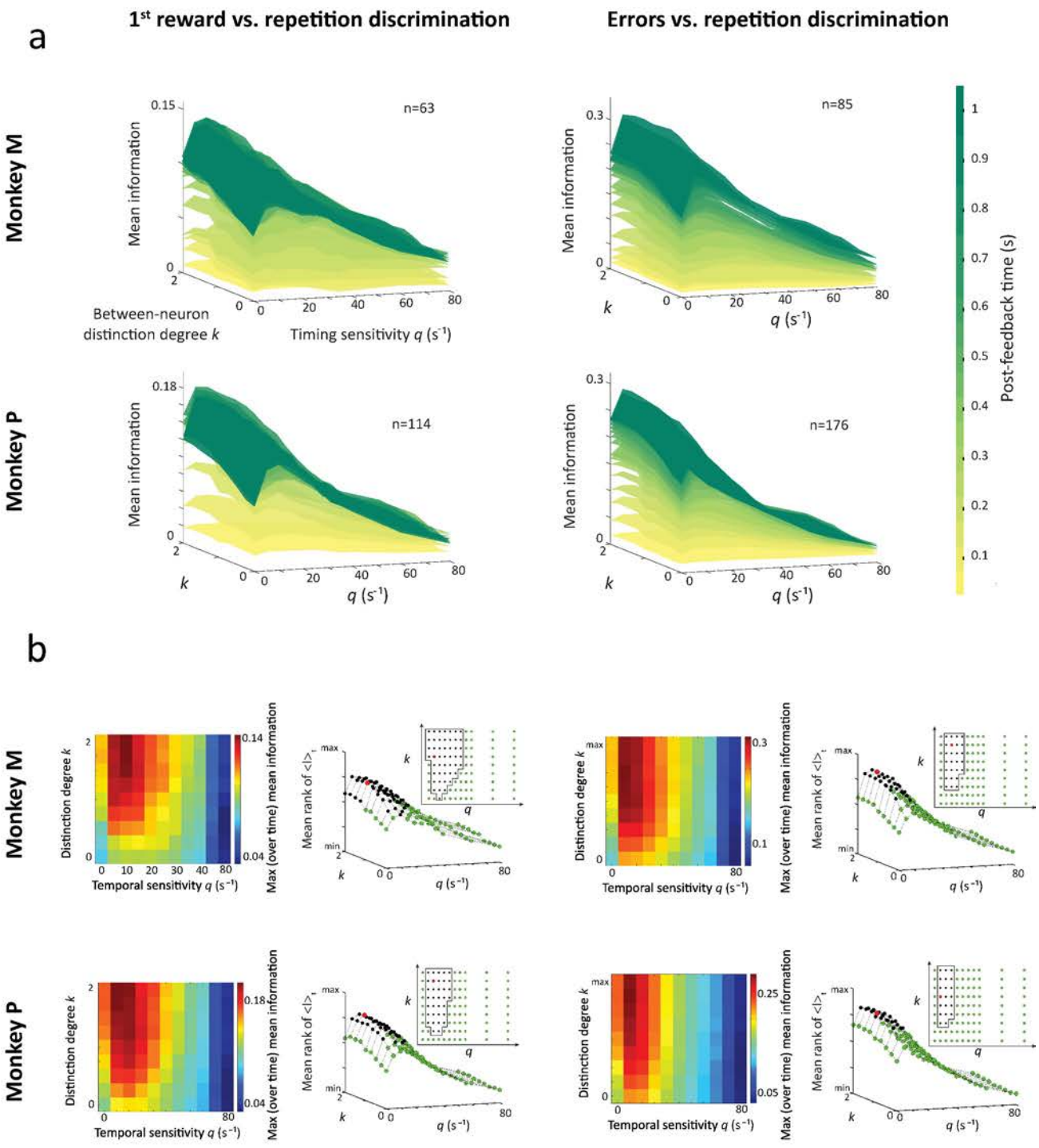


Figure 4.14: *Consistence of the modulation of information in neuron pairs by the temporal sensitivity (q) and the between-unit distinction degree (k) in the two monkeys.* **(a)** Time-course of the mean information among neuron pairs with significant discrimination. Different sheets with different green shadings are different analysis window durations, as indicated on the color scale on the right. For both monkeys and consistently over analysis windows, information increased with adapted temporal sensitivity compared to spike count decoding ($q = 0s^{-1}$), and on average the information was larger for intermediate-to-large values of the discrimination between neurons (k). In this figure, only pairs with significant classification (permutation test) were included, as in Figure 4.12 (b-c). **(b) Left**: maximum (over time-windows) mean (over pairs) information for 1^{st} reward and errors discrimination, as a function of timing sensitivity q and between-unit discrimination degree k. Information was maximized over analysis windows ending in $[0.05, 0.6]$ s, steps of $50ms$, and in $[0.7, 1]$ s, steps of $100ms$. **Right**: comparison of (q, k) for the time-averaged information $\langle I \rangle_t$ of pairs of neurons. The plots display the results of post hoc comparisons (using Tukey's honestly significant criterion correction after a Friedman ANOVA) between $\langle I \rangle_t$ values (computed with analysis windows ending in $[0.1, 1]$ s, steps 100 ms). The red dot marks the $(q, k)_{opt}$ value leading to the higher rank; black dots mark (q, k) values that are not significantly different from $(q, k)_{opt}$, and green dots mark (q, k) values that have significantly smaller ranks than $(q, k)_{opt}$.

4.3.2 Jointly recorded neurons can share similar temporal firing patterns

Decoding with intermediate k values may imply temporal coincidence between spikes from two different neurons as opposed to between spikes from the same neuron. We found that spike coincidence between two neurons occurred, on average, in 34% (1^{st} reward) and 41% (errors) of all pairwise comparisons between spike trains (quantified as during the computation of the between-neuron spike coincidence index in Table 3.3). In addition, we computed an index quantifying the spike coincidence between neurons within a task epoch (defined in Table 3.3). This index negatively correlated with optimal k values, k_{opt_s}, ($c_{1^{st}\,reward} = -0.71$, $c_{errors} = -0.54$, $p < 0.001$). k_{opt} values were pair specific rather than shared among most pairs as for q_{opt}. For instance, k_{opt} values were much smaller for pairs of units that fired preferentially in the same task epoch relative to pairs of units with opposite firing preferences (ranked-sum test, all $p_s < 0.01$; median k_{opt} values were 0.75 vs. 1.25-1.5 for pairs with the same vs. different firing preference). These results suggest that two neurons with similar firing preferences across task epochs were likely to have similar firing temporal patterns.

Some pairs of neurons had maximal $\langle I \rangle_t$ when the decoder did not distinguish between the two single units ($k_{opt} = 0$; 15% of significantly informative pairs, corresponding to 7% and 10% of all analyzed pairs for 1^{st} reward and error discrimination, respectively). These pairs transmitted more information with $q_{opt} \approx 10s^{-1}$ compared to spike count, $q = 0s^{-1}$, (Figure 4.15 (a); signed-rank test on $\langle I \rangle_t$: 1^{st} reward discrimination, $p = 0.029$; error discrimination, $p < 10^{-5}$). They had an index of spike coincidence between neurons larger than in other pairs (Figure 4.15 (b), ranked-sum test: all $p < 10^{-9}$). In these pairs, the information gains relative to the most discriminative unit of the pair were relatively high (Figure 4.15 (c)). This suggested that these pairs were decoded efficiently. We tested whether these information gains were related to the gain of information when not distinguishing between neurons (i.e. $k_{max} = 2$ vs. $k_{opt} = 0$; see definition in Table 3.3). We found a positive correlation (Figure 4.15 (c)), suggesting that spike coincidence between neurons could mediate an efficient combination of their activities.

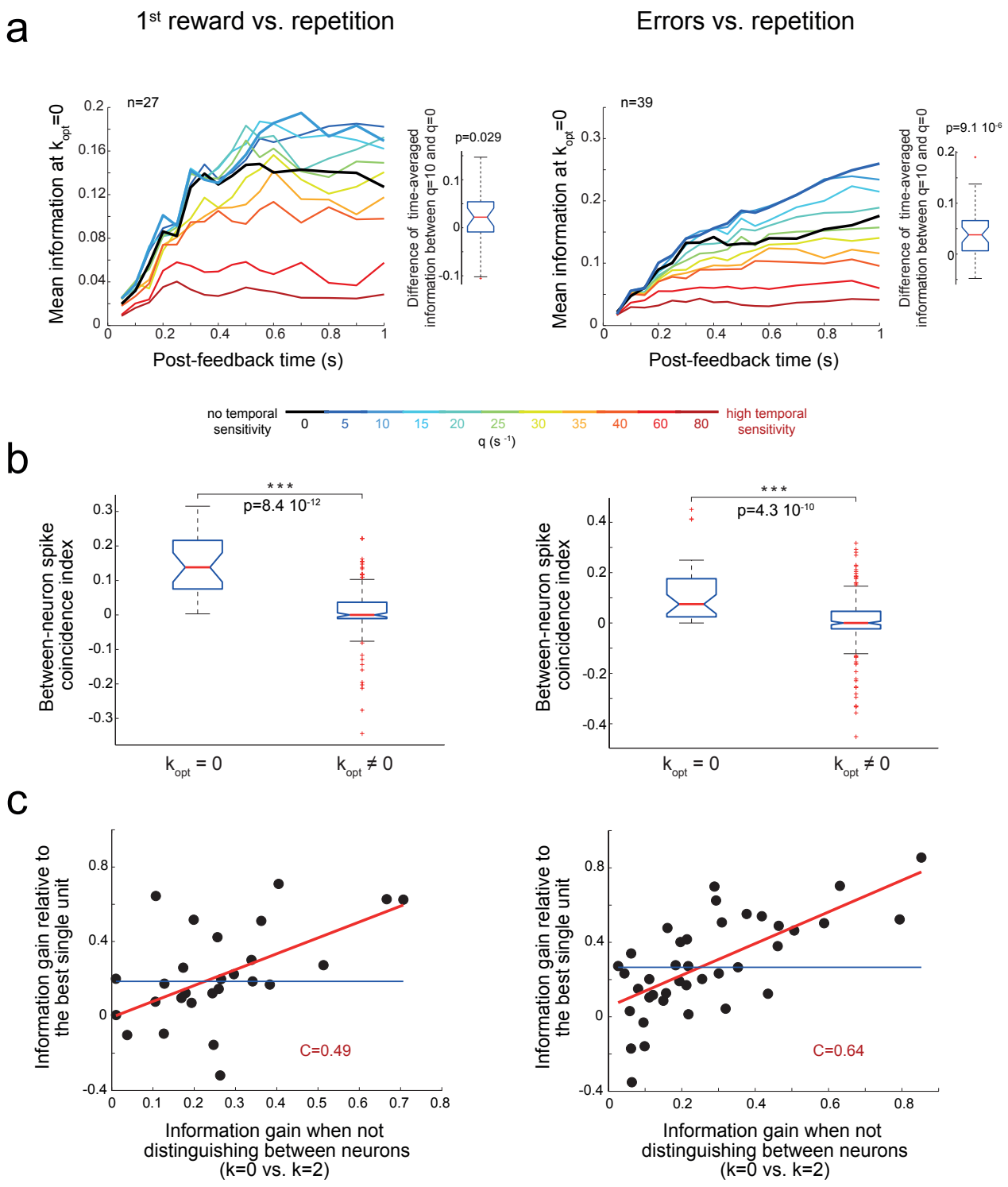


Figure 4.15: *Coding properties of neuron pairs for which* $k_{opt} = 0$. *Left*: Discrimination between the 1^{st} reward and repetition task epochs. *Right*: Discrimination between error and repetition task epochs. **(a)** Left: Mean information among pairs with $k_{opt} = 0$ (significant encoding) as a function of the duration of the analysis window and of temporal sensitivity (q). Right: Distribution of differences in time-averaged information $\langle I \rangle_t$ between $q_{opt} = 10$ and $q = 0s^{-1}$ (for $k_{opt} = 0$). The distribution has a significantly positive median (signed rank test). **(b)** The index of spike coincidence between neurons was higher for pairs with $k_{opt} = 0$ compared to other significant pairs (ranked-sum test, $p < 10^{-9}$). Note that the median indexes were larger than 0 for pairs with $k_{opt} = 0$. This means that when comparing spike trains within one task epoch, coincidences between neurons occurred more often than when comparing spike trains between task epochs (see the definition of this index in Table 3.3). **(c)** The information gain relative to the most informative single unit was positively correlated with the information gain induced by the absence of neuron distinction. C: Spearman correlation coefficient, red line: and linear fit, blue line: median of the distribution of information gains.

Hence, on the one hand the information generally increased when the identity of the neurons was accounted for (intermediate-to-high k_{opt} values), which indicates that reliable spike times were variable and distributed across the neuronal population. On the other hand, some pairs of neurons with similar temporal firing patterns could be efficiently decoded through between-neuron temporal coincidences (Figure 2.1 (c), Figure 3.1 (b)).

4.4 The temporal structure of single unit spike trains predicts behavioral response times

The presence of information in single-unit spike timing does not necessarily imply that the downstream networks do actually use it [Luna et al. (2005); Carney et al. (2014)]. In particular, if dACC spike timing were not used, then different temporal patterns would be rather unlikely to reliably correlate with different behavioral outputs. Here we examined whether 1^{st} reward single-unit activity could predict upcoming behavior. We focused on the behavioral response time, i.e. the time between the GO signal (for hand touch) and the following touch on target (section 3.5). The response time was measured during the trial following the 1^{st} reward, i.e. several seconds after the analyzed neural activity. This behavioral response time was quantifying how long it took to the monkey to confirm its choice (after saccading), during what should be the beginning of the repetition period (unless a mistake was made, which happened in less than 2% of the trials).

The modulation of the response times of the monkeys was rather consistent with a relation to cognitive control, rather than with a purely motor effect. Indeed:

1. The time taken by the monkey to release the central lever (which was an identical movement for all targets) and the response time were similarly modulated (see Figure 4.16 (a-b)).

2. While the two monkeys were in the same apparatus, the modulation of the response time by the target was monkey-specific (see Figure 4.16 (a-b)). This rather points towards a spatial attention effect, which would be consistent with the fact that monkeys were more likely to begin a problem by touching a specific target [Khamassi et al. (2014)].

3. The response times of both monkeys consistently increased on the touch following the 1^{st} reward compared to the touch leading to 1^{st} reward

(Figure 4.16 (c)). This was in agreement with a behavioral switch from exploration to repetition.

4. The long response time trials were associated with a larger probability of (preceding) interruption of the trial, due to breaks in fixation or touch requirement (Table 4.1).

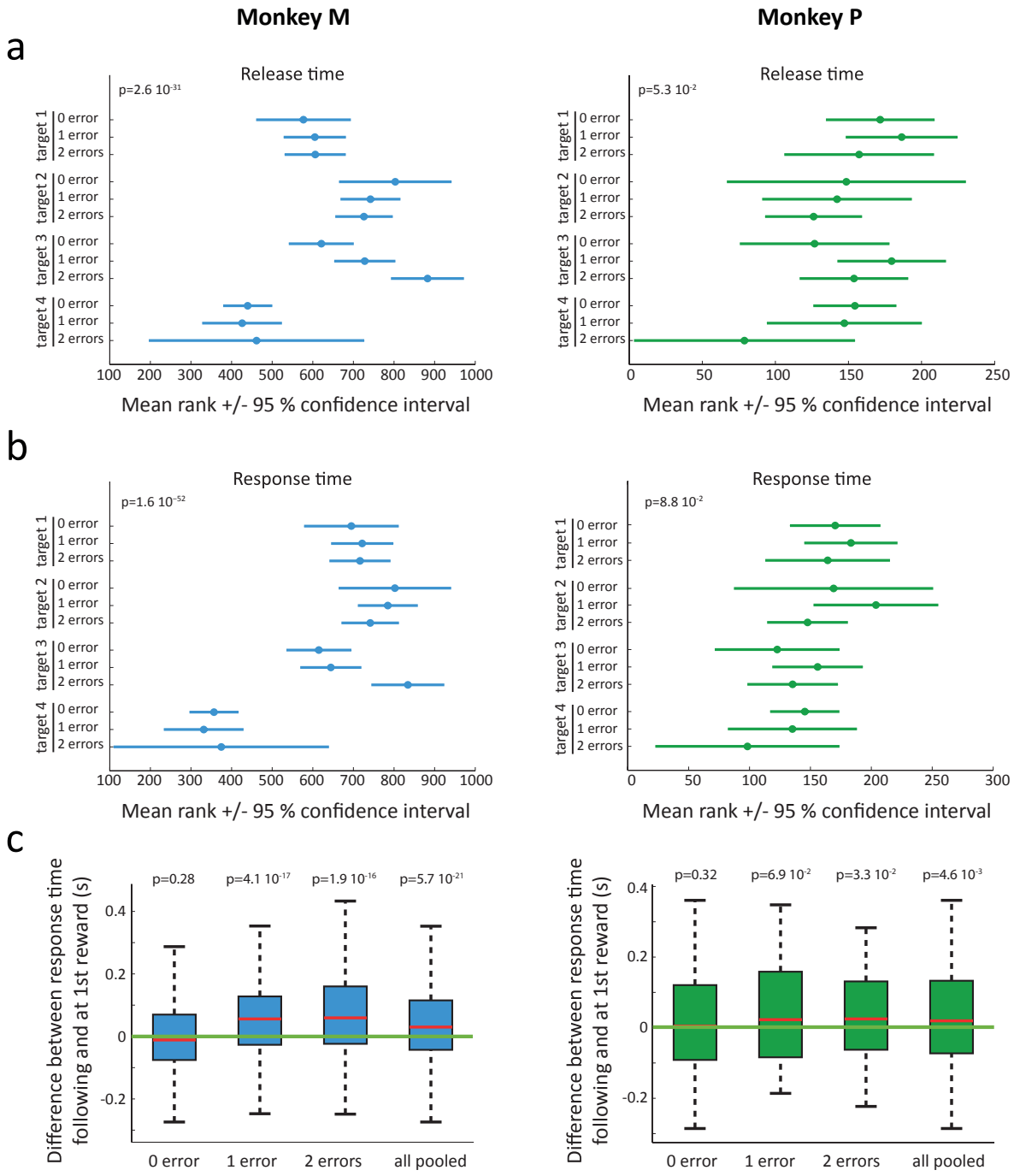


Figure 4.16: *Modulation of behavioral response times following 1^{st} reward trials.* The analysis was restricted to the trials that are used for the analysis linking neural activity to future behavior (in Figure 4.17). **(a)** Modulation of the release time following 1^{st} reward by the identity of the rewarding target and by the number of errors made preceding the 1^{st} reward. The release time was defined as the time between the post-1^{st}-reward go signal for target touch (by the hand) and the release of the central lever button. Groups were compared with a non parametric Kruskal-Wallis test (see p-value at the top-left). Post-hoc comparisons were conducted using Tukey's honestly significant criterion correction. Note that for all rewarding targets, the release movement occured at the same place: on the central lever button. The release time modulation is therefore not likely to reflect motor constraints. **(b)** Modulation of the response time following 1^{st} reward by the identity of the target and by the number of errors, conventions as in a). The response time was defined as the time between the post-1^{st}-reward go signal for target touch and the following target touch. The modulation of the response time was strikingly similar to the modulation of the release time (which, as argued above, is very unlikely to reflect motor constraints). In addition, note that while the two monkeys were in the same apparatus, they modulated their response time differently for the different targets. Finally, the target modulation of response time could interact with the modulation by the number of preceding errors. Altogether, the results argue against a purely motor cause for response time modulation, and rather point toward a spatial bias of cognitive processes. **(c)** Boxplots for the difference of response time between the trial following 1^{st} reward (the 1^{st} repetition, or, in rare cases, a mistake) and the trial that ended with the 1^{st} reward, i.e. last exploration. The p-value of a signed rank test for a bias of the distribution toward either positive or negative values is indicated. The green line indicates a 0 difference. For clarity, outliers are omitted. The response time increased on the trial following 1^{st} reward when the preceding exploration period was longer than one attempt.

	Monkey M	Monkey P	Both monkeys
Difference of mean number of aborted trials	Median=0.19 $p_{signrank}$=5.3 10^{-4}	Median=0.095 $p_{signrank}$=0.15	Median=0.16 $p_{signrank}$=3.2 10^{-4}
Difference of probability of errors	Median=0, mean=-9.2 10^{-3} $p_{signrank}$=0.31	Median=0, mean=9.7 10^{-4} $p_{signrank}$=1	Median=0, mean=-5.9 10^{-3} $p_{signrank}$=0.36

Table 4.1: Difference of probability of mistakes or of mean number of trial interruptions after 1^{st} reward, between problems with response time larger than median and problems with response time lower than median (response time measured between the first post-1^{st}-reward go signal and the post-1^{st}-reward touch). These interruptions can be due to break of fixation or break in screen touch requirements, after which monkeys were forced to resume the trial (see section 3.1). The medians (and means for differences in mistake probability) of these differences are shown together with a signed rank test measuring how significantly the median deviates from 0. Note that the overall percentage of mistakes was very small (0.81% and 1.0% in monkey M and monkey P, respectively, of considered trials).

We separated trials into two groups: one group with response times larger than the median, and the other with response times below the median. The probability of switching to repetition was very high in both groups and statistically equivalent between them (Table 4.1). We tested the hypothesis that longer response times may reflect a longer decision-making process, when monkeys might act more carefully to avoid mistakes.

4.4.1 Deviations from prototypical temporal firing patterns predict response times

Under the temporal decoder hypothesis that we suggested (Figure 2.1 (c)), the success of information transmission relies on matching the discharge received during a particular trial with a prototypical activity pattern specific to a given task epoch. The robust classification of single-unit spike trains (Figure 4.2, Figure 4.8) indeed implies that during many 1^{st} reward trials, the activity resembled a prototypical firing pattern specific to a feedback triggering behavioral adaptation. However, the classification was not perfect, which suggests that there were also trials during which the activity deviated from the prototypical temporal firing pattern. This could lead to inefficient information transmission, and then slower processing.

To test this, we developed a new method to estimate how much each

single-trial spike train emitted by each neuron at 1^{st} reward deviated from its "prototypical" discharge (i.e. its more common firing pattern during 1^{st} reward; see section 3.5). According to our method, for each neuron, a large positive deviation from prototype can occur when the spike count is either higher or lower than the average rate of the neuron, or when the spike times are jittered compared to the neuron's usual temporal pattern. Hence, when computing the deviation based on $q = 0$, high values of deviation will in general be attributed to trials with both large and small spike count relative to average (as long as the spike count distribution is not overly skewed, with outliers lying on one side only). We compared the two groups of trials: associated with slow vs. fast response times. For each neuron, we computed the difference in mean deviation from prototypical activity ($\overline{D}$) between these two groups. Notably, the distribution of $\overline{D}$ values was positively skewed (Figure 4.17 (a,c)), indicating that a larger deviation from prototypical activity predicted a longer behavioral response time. This effect was consistent in both monkeys and between different subpopulations of neurons (Figure 4.18).

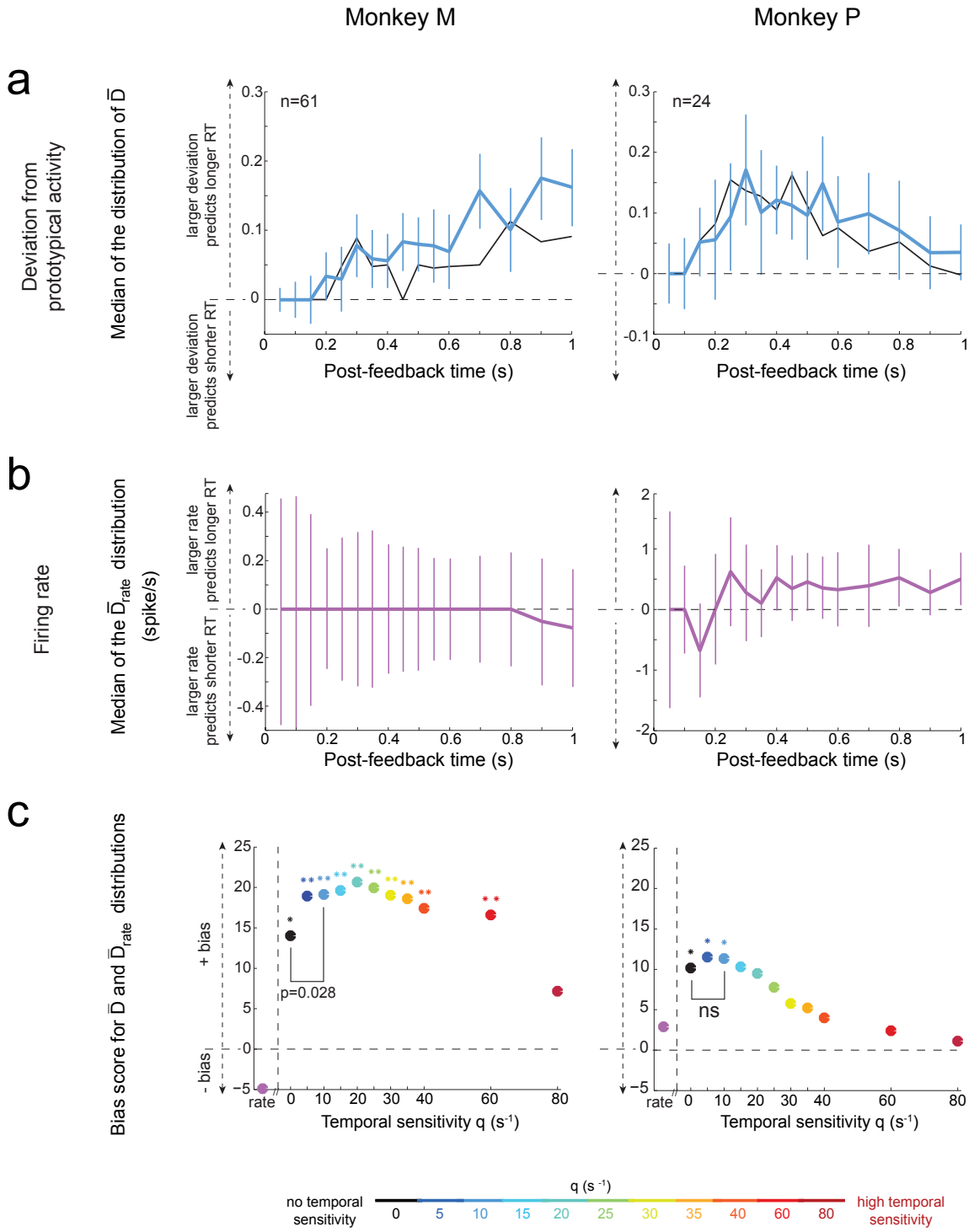


Figure 4.17: *The temporal structure of single unit spike trains predicts behavioral response times. Left*: Monkey M (all significantly informative neurons for 1^{st} reward vs. repetitions). *Right*: Monkey P (significant neurons with information $\geq$ median; neurons with very little information did not permit robust behavioral prediction in this monkey, see Figure 4.18). Analysis windows end at the time indicated by the x-axis. **(a)** Test for the spatiotemporal decoder. Time course of the median $\overline{D}$. The value of $\overline{D}$ is positive when spike trains emitted in 1^{st} reward trials followed by slower response times deviate more from prototypical spike trains than those emitted in trials followed by fast response times. The two curves correspond to $q = 0$ (black) and $q = 10s^{-1}$ (blue). Bars represent a median confidence interval (see section 3.6 for the definition). **(b)** Test for the neural integrator decoder receiving excitatory inputs from dACC feedback-related neurons. Time course of the median $\overline{D}_{rate}$ (difference in mean firing rate between trials with slow and fast response times). The value of $\overline{D}_{rate}$ is positive if trials with high rates tend to be followed by long response times. **(c)** Bias scores (across different analysis windows) for $\overline{D}$ and $\overline{D}_{rate}$. A large positive bias score indicates that the data is very positively skewed (relative to a distribution that is symmetrically distributed around 0). Stars indicate significance values for these biases (2-sided permutation test: *, $p \leq 0.05$; **, $p \leq 0.01$). For Monkey M, the lowest p-value was for $q = 20s^{-1}$ ($p = 0.003$); for Monkey P, the lowest p-value was for $q = 5s^{-1}$ ($p = 0.029$). Finally, the result of the comparison of $\overline{D}$, averaged over different analysis windows, between $q_{opt} \approx 10s^{-1}$ and $q = 0s^{-1}$ is shown (signed-rank test).

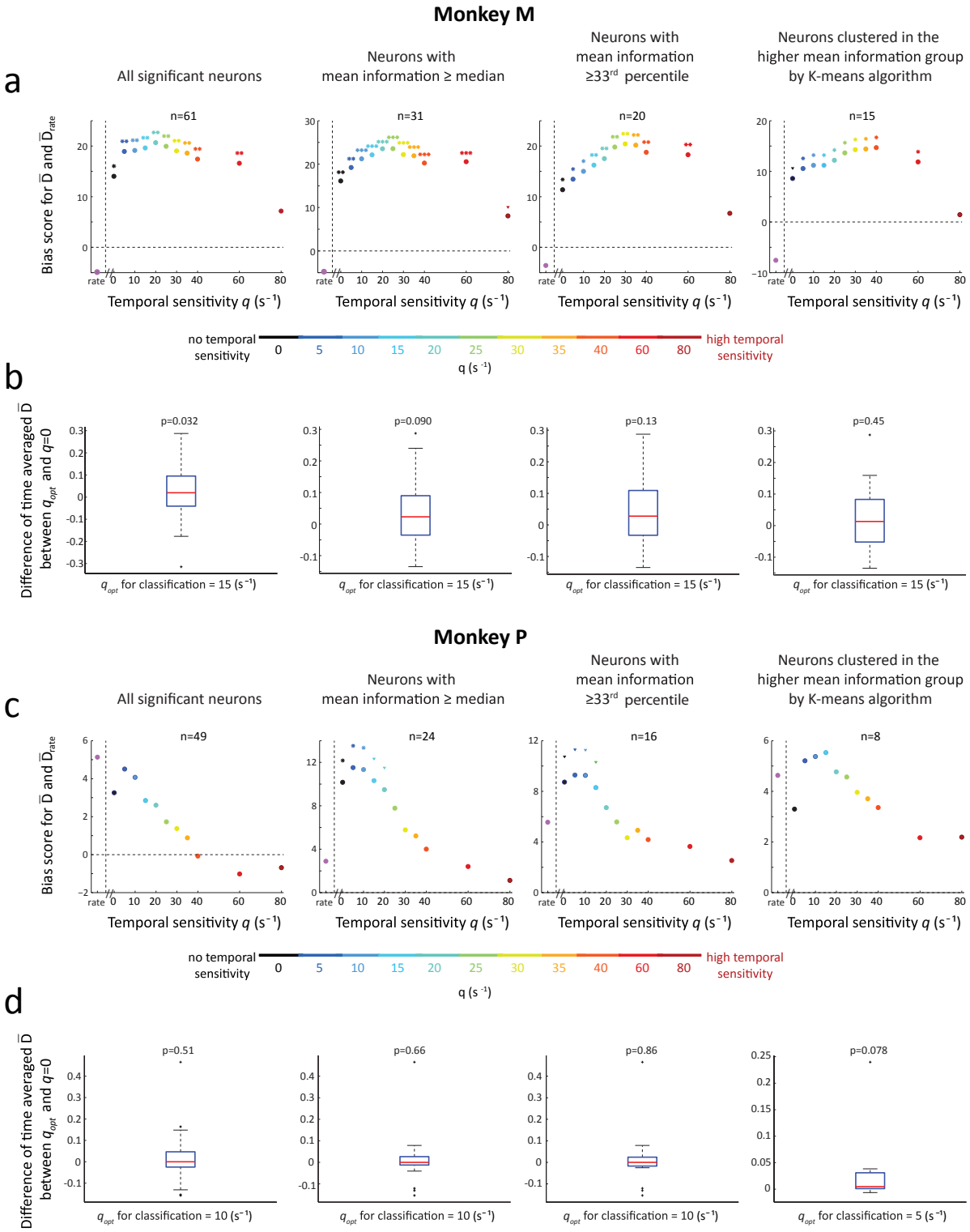


Figure 4.18: *Consistency of the relation between neural activity and behavior in different subgroups of neurons.* Longer response times were observed in trials preceded by larger deviations from prototypical spike train (i.e. $\overline{D} > 0$), consistently for different subgroups of neurons with significant 1^{st} reward *vs.* repetition classification. After this classification, we ranked neurons according to $I_{max} = \max_q(\langle I \rangle_t)$ (information computed using the original Victor and Purpura metric). We formed different subgroups more and more restricted to high information neurons, as indicated. The smallest group was formed by applying a k-means algorithm (2 clusters) and taking only the high-information cluster. **(a, b)**: monkey M, **(c, d)**: monkey P. **(a,c)** Bias score for $\overline{D}$ and $\overline{D}_{rate}$ as a function of the set of considered neurons. Neurons with less than 5 available trials were discarded. The p-value (2-sided permutation test) is indicated for each data point: small triangle for $p <= 0.1$; one star for $p <= 0.05$; two stars for $p <= 0.01$; three stars for $p <= 0.001$. Note that the values of $\overline{D}_{rate}$ in this figure are computed as in Figure 4.17 (they assume positive weighting of all neurons). **(b,d)** Comparison of $\overline{D}$ values between q_{opt} and $q = 0$. Here, q_{opt} was the temporal sensitivity that maximized discrimination between 1^{st} reward and repetition using the normalized distance d^* in each neuronal group (see section 3.5). Note that similar results were found when using $q_{opt} = 10s^{-1}$ instead, i.e. the temporal sensitivity that maximized 1^{st} reward discrimination when using the original Victor and Purpura distance (as in Figure 4.17 (c)). $\overline{D}$ was time-averaged (over analysis windows ending in $[0.1, 1]$ s, steps of 100 ms), separately for q_{opt} and $q = 0$. These time-averages were compared with a signed rank test (p-value indicated). The boxplots represent the distribution of the difference of time-averaged $\overline{D}$ between q_{opt} and $q = 0$.

For monkey P, statistical robustness was reached when neurons with very little 1^{st} reward vs. repetition information $\langle I \rangle_t$ were removed (Figure 4.18 (c)). Temporal sensitivity values leading to best task-epochs decoding (Figure 4.2) were also relevant for predicting behavioral response times. More precisely, q values of 5 and 10 s^{-1} led to a robust and significant bias of the distribution of $\overline{D}$ values in both monkeys (Figure 4.17 (c); and Figure 4.18). This result suggests that in both monkeys the temporal patterns of spikes could be relevant to downstream decoding areas ultimately adapting behavioral responses. We note that the deviation from prototype based on spike count ($\overline{D}(q = 0)$) was also significantly biased in both monkeys. Importantly, we confirmed that this effect was not likely to be merely caused by a difference in firing rate between trials with slow and fast response times with different signs in different neurons. Indeed, if the latter hypothesis were true, then the neurons with a large $\overline{D}(q = 0)$ would also be those with large absolute value of the difference of mean firing rate $\overline{D}_{rate}$ between slow and fast response time trials. This would lead to a strong correlation between $\overline{D}(q = 0)$ and $\overline{D}_{rate}$. Instead, we found that this correlation was small (Spearman correlation between time-averaged $\overline{D}(q = 0)$ and $\overline{D}_{rate}$, monkey M: $c = 0.31$, monkey P: $c = 0.06$). This suggests that, for any single unit showing a large $\overline{D}(q = 0)$ and a small $\overline{D}_{rate}$, the trials followed by long response times were likely to be associated with both increased (during some trials) and decreased (during other trials) spike count compared to prototype.

We then compared the deviations from prototype between the method based on spike count ($q = 0$), and the method using the temporal sensitivity $q_{opt} \approx 10s^{-1}$ (optimal to discriminate 1^{st} reward vs. repetitions). For monkey P, we found that spike count and temporal decoding performed equally well (signed-rank test on average $\overline{D}$ over analysis windows ending from 0.1 to 1s by increments of 0.1 s). However, the two decoding strategies probably relied on different neurons as $\overline{D}$ values were considerably different between q_{opt} and q=0. Indeed, we tested whether the difference $Diff_{\overline{D}} = \overline{D}(q = 10) - \overline{D}(q = 0)$ was likely to act as a negligible noise (that was thus exchangeable between neurons) for the bias score of $\overline{D}(q = 10) = Diff_{\overline{D}} + \overline{D}(q = 0)$. However, this had a very small probability to happen: 2% (this figure is the p-value of a permutation test, see subsection 3.5.3 for the method). In monkey M, optimal temporal sensitivity significantly improved the relation between single-unit spike trains and upcoming response times compared to spike count ($p = 0.028$; Figure 4.17 (c); see also Figure 4.5 (d)). Altogether, these results further argue in favor of the

relevance of temporal spiking patterns for behavioral adaptation.

Note that even though a longer response time was associated with a higher probability of interruptions in the task (e.g., breaks in fixation) during the trial ending with this response (Table 4.1), the correlation between response time and neural activity was not entirely caused by a difference between interrupted vs. uninterrupted trials. Indeed, when we removed interrupted trials we still observed a significant positively skewed $\overline{D}$ distribution (Figure 4.19).

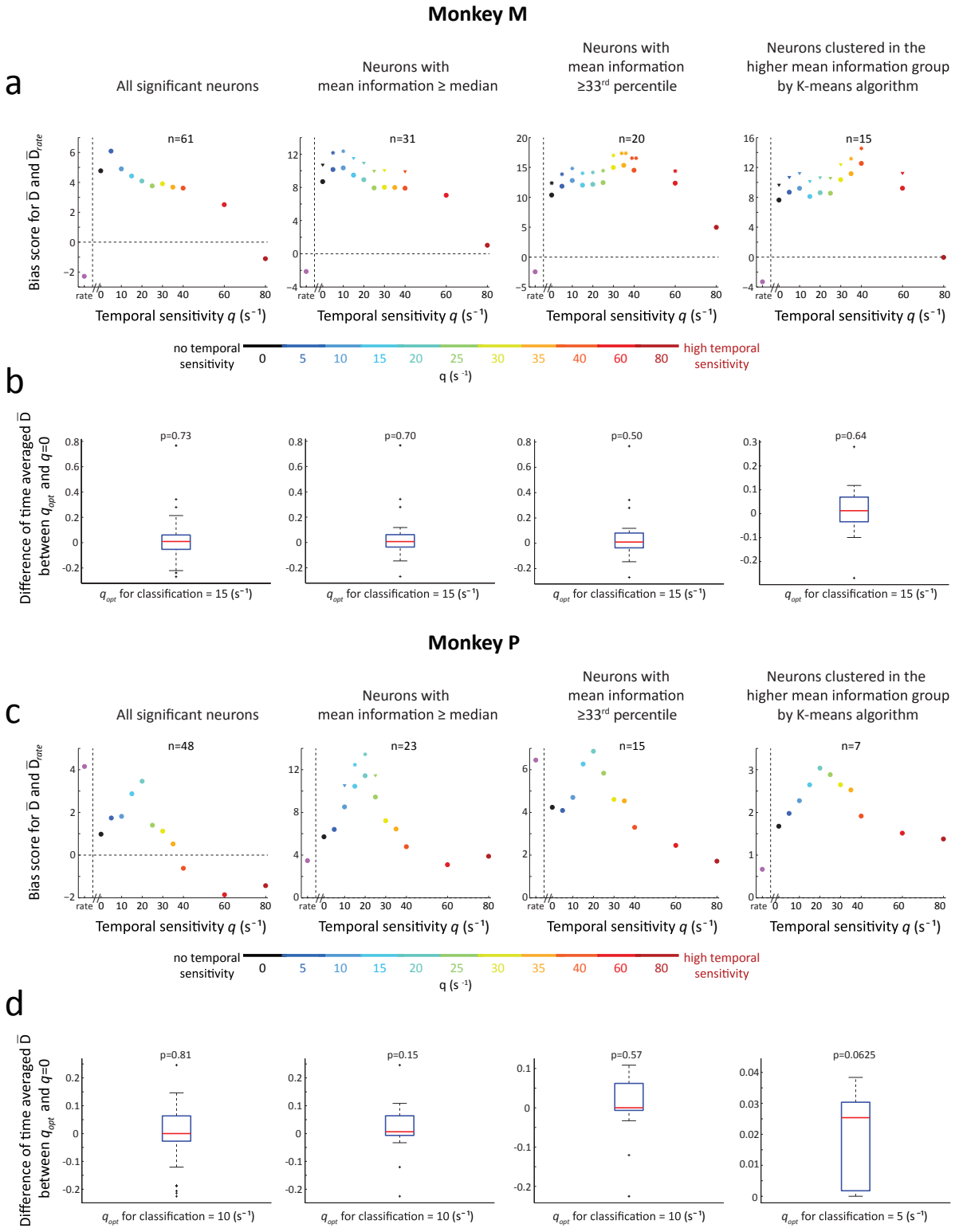


Figure 4.19: *The relation between neural activity and behavior was still present when excluding trials with interruptions.* Behavioral response time analysis while excluding post-1^{st}-reward trials that were interrupted before the monkey touches the target. These interruptions can be due to breaks of fixation or breaks in screen touch requirements, after which monkeys were forced to resume the sequence of actions (section 3.1). Groups of neurons are as in Figure 4.18. **(a, b)**: monkey M, **(c, d)**: monkey P. **(a,c)** Bias score for $\overline{D}$ and $\overline{D}_{rate}$ as a function of the set of considered neurons. Neurons with less than 5 available trials were discarded. The p-value (2-sided permutation test) is indicated for each data point by the following symbols: small triangle for $p <= 0.1$; one star for $p <= 0.05$; two stars for $p <= 0.01$; three stars for $p <= 0.001$. Note that the values of $\overline{D}_{rate}$ in this figure are computed as in Figure 4.17 (they assume positive weighting of all neurons). **(b,d)** Comparison of $\overline{D}$ values between q_{opt} and $q = 0$. Here, q_{opt} was the temporal sensitivity that maximized discrimination between 1^{st} reward and repetition using the normalized distance d^* in each neuronal group (see section 3.5). Note that similar results were found when using $q_{opt} = 10s^{-1}$ instead, i.e. the temporal sensitivity that maximized 1^{st} reward discrimination when using the original Victor and Purpura distance (as in Figure 4.17 (c)). The values of $\overline{D}$ were time-averaged (over analysis windows ending in $[0.1, 1]$ s, steps of 100 ms), separately for q_{opt} and $q = 0$. The resulting time-averages were compared with a signed rank test (p-value indicated). The boxplots represent the distribution of the difference of time-averaged $\overline{D}$ between q_{opt} and $q = 0$.

Note also that our results could not be explained by a segregation of 1^{st} reward responses into 4 equidistant clusters corresponding to the 4 targets. Indeed, under this hypothesis, all spike trains should have had similar values of neural deviation from prototypical activity, as the latter is averaged over all spike trains associated to all targets. Therefore, in this case, Figure 4.17 should not show significant differences of deviation from prototypical activity between different groups of spike trains. This suggests that dACC activity was not merely related to movement coding. Rather, our results indicate that the deviation of dACC activity from prototypical temporal patterns could mediate a behavioral adaptation process modulating the delay of upcoming decisions and actions.

4.4.2 Firing rate increase does not robustly relate to a behavioral response time change

In the context of a neural integrator decoder maintaining a memory of the necessity to adapt the behavioral strategy, one could expect that the spike count would be directly predictive of the behavioral response time. Indeed, in this scenario, the downstream decoder would receive an overall excitatory input from the population of dACC neurons whose activity distinguishes between 1^{st} reward and repetition task epochs, as this population fires more on average during 1^{st} rewards (Figure 4.8). As a consequence, any decrease in the number of spikes received by the decoder would hinder reaching the decision-making threshold (see "adapt behavioral strategy threshold" in Figure 2.1 (c)). Conversely, any increase in spike input would accelerate threshold crossing. Hence, given the two groups of trials (slow vs. fast response times), we tested whether dACC neurons fired more in one of these two groups. We computed the difference in mean firing rate between spike trains that preceded trials with slow vs. fast response times ($\overline{D}_{rate}$, see section 3.5). We found that the distribution of $\overline{D}_{rate}$ was not significantly skewed either positively or negatively, indicating that large firing rates in dACC neurons were not predictive of future monkey's response times (Figure 4.17 (b,c), see also Figure 4.18).

In addition, under the neural integrator decoding hypothesis, the dACC neurons firing more during 1^{st} reward (70% of significant neurons) are expected to be the main drivers of firing rate increase in the decoder. To examine this, we restricted the analysis to neurons firing more during 1 s after the 1^{st} reward (compared to repetitions). We found that this restriction did not lead to a more

robust bias of the of $\overline{D}_{rate}$ distribution (2-sided permutation test: monkey M, $n = 42$, bias score $= -5.50$, $p = 0.28$; monkey P, $n = 18$, bias score $= -0.337$, $p = 0.93$). The same test made on different subgroups of neurons (the subgroups described in Figure 4.18) also failed to reach significance (all $p_s > 0.05$).

Finally, we also examined a scenario in which the downstream integrator decoder would receive excitatory inputs from neurons discharging more during 1^{st} reward, and inhibitory inputs from neurons discharging more during repetition. We simply reversed the sign of $\overline{D}_{rate}$ for those neurons discharging more during repetition. However, using the same neurons as for Figure 4.17, we did not find a robust bias of the overall resulting distribution (2-sided permutation test: monkey M, $n = 61$, bias score $= -4.14$, $p = 0.37$; monkey P, $n = 24$, bias score $= -3.0$, $p = 0.48$). The same test was made on different subgroups of neurons (the groups in Figure 4.18). The absolute value of the rate bias score reached by using this methodology was never higher than the corresponding bias score reached by using the best-scoring measure of deviation from prototypical spike train. Furthermore, this rate bias score reached $p < 0.1$ only once, for the smallest group of neurons of monkey M (bias score -11.8, $p = 0.037$). We note that this smallest group of neurons was not the one associated with the largest effect size for behavioral prediction through deviation from prototypical pattern at q_{opt} (the median time-averaged $\overline{D}(q = 10s^{-1})$ was of 0.11 in this smallest group of neurons, while it was of 0.19 in the group of neurons with information larger than the median in this monkey). Note also that this effect in the smallest group of monkey M neurons was not statistically very robust (compare to the much smaller p-values reached when using $\overline{D}$ in Figure 4.18 (a)); it relied on few neurons, with less than 15 trials available for one third of them. We stress that if one undersamples the spike count variability, there is some non-negligible probability that, by chance, one mostly samples outliers on one side of the distribution. This could lead to a situation where a measure of absolute deviation from prototype and a measure of spike count difference can covary, hence being difficult to distinguish (as it seems to happen here). Furthermore, when using this methodology for computing $\overline{D}_{rate}$, we found rate bias scores that could appear inconsistent between monkeys. Indeed, in contrast to the negative rate bias scores of monkey M, for monkey P this rate bias score was – non-significantly – positive for the smallest group of neurons (when computed with or without the trials with interruptions). Finally, the rate bias scores computed by using this methodology never reached significance when considering only trials without interruption (as in Figure 4.19,

all $p_s > 0.05$).

Overall, the results suggest that there was no robust monotonous relation between the firing rates of dACC feedback-related neurons, and the behavioral response time changes.

These observations therefore appear hard to reconcile with the hypothesis of a decoding by a simple downstream integrator.

In contrast, the robust relation between deviations from a neuron-specific prototypical 1^{st} reward spike train and slower upcoming response times could be consistent with a non-linear downstream network able to process and separate different spatiotemporal spiking patterns.

Chapter 5

Discussion: evidence for a temporally sensitive, non-linear decoder of dorsal Anterior Cingulate Cortex signals

---

Post-feedback spike counts in dACC neurons were shown to depend on whether behavioral adaptation was required [Quilodran et al. (2008)]. Given the absence of external-stimulus-driven temporal fluctuations in the synaptic input and high noise in spike timing, a plausible hypothesis would be that only spike count is relevant to the transmission of the need to adapt behavior by dACC firing [London et al. (2010)].

5.1 Evidence for internally generated reliable temporal structure and spike count variability in dACC

By contrast, we provide evidence for an efficient spatiotemporal spike coding of behavioral adaptation signals. Our analysis accounts for the temporal sensitivity of a biologically plausible neural decoder which would receive post-feedback dACC discharges. Adjusting the temporal sensitivity of the decoder can enhance the readout of single-unit spike trains relevant to behavioral adaptation. Beyond the existence of a temporal patterning of dACC activity, these results indicate that spike-timing reliability supplements spike-count reliability. Interestingly, in frontal areas, single-unit spike generation mechanisms or network dynamics, rather than external stimuli or motor

feedback, are probably responsible for spike timing reliability and spike-count variability [Litwin-Kumar and Doiron (2012); Mongillo et al. (2012); Pozzorini et al. (2013); Ostojic (2014)]. We found strong temporal correlations, stronger-than-Poisson spike count variability, and heterogeneous spike times across the dACC population. The feedback-type-specific spiking dynamics of dACC is thus unlikely to arise from neuronal populations connected by balanced excitatory and inhibitory inputs with uniform wiring probability and with stationary weak-to-moderate strengths [Litwin-Kumar and Doiron (2012); Ostojic (2014)], as these features would tend to create Poisson-like spike trains. Besides the effect of the network's connectivity pattern, spike-triggered hyperpolarizing currents or short-term plasticity could also plausibly favor the presence of informative temporal correlations in dACC activity [Arsiero et al. (2007); Mongillo et al. (2012); Farkhooi et al. (2013)]. In addition, spike-triggered hyperpolarizing currents (i.e., single-neuron adaptation) may participate to shaping the lower-than-Poisson spike count variability occurring shortly after the feedback [Farkhooi et al. (2011)]. This initial small spike-count variability also suggests that the synaptic current received by the neurons just after the feedback could be characterized by relatively small fluctuations [Litwin-Kumar and Doiron (2012); Schwalger and Lindner (2013)].

Note that the optimal range of decoding time scale that we found ($\tau \approx 70 - 200ms$) is larger than those found when decoding responses to stimuli with relevant temporal patterning or contrast at onset time (e.g., auditory stimuli, $\tau \approx 5ms$ [Machens et al. (2003)]; visual stimuli, $\tau \approx 10 - 100ms$ [Victor and Purpura (1996); Aronov et al. (2003)]). This is consistent with the idea of a hierarchy of increasing time scales from sensory to higher-order areas [Murray et al. (2014)]. However, there are also exceptions to this rule, for instance in the gustatory modality (for which the timing of the stimulus is less relevant). Indeed, the optimal time scales were found to be close to the one we found in dACC (50-500 ms [Roussin et al. (2012)]). Given that during a gustatory stimulation, the motor behavior of the animals and/or some sensorial input transients were probably participating to shaping the temporal code [Roussin et al. (2012)], it is quite remarkable that we found equivalent time scales in our data for which internal neuronal dynamics was probably the major contributor to spike timing reliability. From a functional view-point, in our context, a time-scale of $\approx 70 - 200ms$ may be considered as short for two reasons. First, it is shorter that the time-interval during which subpopulations of dACC neurons, or even single dACC units, appear to increase their firing rate during the

feedback. Second, and perhaps more importantly, such a time-scale would not permit to maintain the memory of the behavioral adaptation signal through leaky integration. Indeed, this suggests a weakness of a downstream network that would implement, as a decoding and memory mechanism, a computation tantamount to an approximate integration. Such a network would probably be less robust than a spike-timing sensitive downstream decoder.

We note that the optimal spike coincidence timescale loosely matches the period of local field potential (LFP) oscillations in the delta and theta range, on which frontal neurons can phase lock during cognitive tasks [Benchenane et al. (2010); Womelsdorf et al. (2010); Totah et al. (2013); Womelsdorf et al. (2014)]. LFPs partially reflect the synaptic input of the local population [Reimann et al. (2013)], which could both shape and be influenced by the temporal spiking patterns of dACC. The optimal temporal sensitivity range for decoding identified in this study remains an approximation. First, different methods or different analysis windows might give slightly different optimal values (Figure 4.2, Figure 4.3). Yet, although it is not feasible to extensively test all possible decoders, our analysis accounts for biophysically reasonable assumptions on the downstream decoder. In this framework, we provide strong evidence for the plausibility of decoding through spike coincidences (up to a few hundred ms), compared to a neural integrator decoder. Second, spike trains were referenced to feedback time, but the internal reference of the brain could be different and more or less accurate [Chase and Young (2007)] (e.g., coincidence detection during a population onset [Panzeri et al. (2010)], or precise spike timing relations in a neuronal population [Shmiel et al. (2006)]). Aligning to feedback times was very relevant for behavioral-adaptation task epochs where monkeys could not predict the outcome and were thus reacting to feedback. However, anticipation of rewards during repetition periods may have promoted internal references dissociated or jittered from actual juice delivery, decreasing the apparent temporal reliability (as suggested by the data, Figure 4.8).

5.2 A biological architecture could decode dACC temporal signals

The spike-time sensitive decoder can be understood as a downstream network that, through synaptic plasticity [Gjorgjieva et al. (2011)], becomes differentially

selective to coincident spiking patterns that are specific to task epochs. The optimal temporal sensitivity range is compatible with the time constant of NMDA-mediated currents. Indeed, the efficiency of the spike coincidence mechanism decreased with interspike intervals up to 200 ms, which relates to an exponential decay time-constant of 100 ms.

Within this framework, decoding thus relies on the convergence of excitatory neurons that transmit similar temporal patterns to a post-synaptic compartment (triggering summation of depolarizations). Yet, informative neurons with distinct and potentially antagonistic temporal patterns may improve information transfer, for instance if they were decoded by different specialized post-synaptic neurons. We showed that paired decoding generally enhanced information transmission relative to the pair's most discriminative unit. This suggests that highly informative activity can be advantageously combined (the less informative inputs do not merely act as contaminating noise on average). The information increase was achieved by varying the degree of distinction between the two units (parameter k). This mechanism may be implemented by different spatial organizations of synapses, which could modulate, through non-linear summation, the temporal precision of spike coincidence detection. Other mechanisms such as different synaptic weights or synaptic timescales (i.e. two weak/shorter depolarizations that require more precise coincidence to efficiently sum), or targeted inhibition, may also induce a similar effect. In addition, we showed that in a smaller proportion of pairs the activity of both units did not need to be distinguished to achieve optimal discrimination. Thus, if these two units were excitatory, direct summation of their post-synaptic potentials would be advantageous. The partial spatial specificity of reliable spikes may be advantageous during realistic decision-making when quick choices should be made between many strategies. Indeed, the combination of spatial and temporal information can increase the number of possible specific activity patterns compared to simultaneous firing of all neurons.

5.3 Evidence for a relation between future behavior and the result of a non-linear, spike timing sensitive decoding of dACC signals

We further probed dACC function by testing how it could affect future behavior. We found a significant correlation between neural activity at feedback time and the monkeys' response time during the following trial. This finding is functionally different from the correlation previously reported between pre-movement dACC activity (which often resembles an integration to threshold [Hayden et al. (2011b); Michelet et al. (2015)], in contrast to feedback-driven dACC responses) and immediate motor response [Hayden et al. (2011b); Sheth et al. (2012); Michelet et al. (2015)]. This motor correlation could become apparent through the comparison between trials with high vs. low firing rates (or, equivalently, spike-counts in a given window). In particular, Michelet et al. showed that the quicker the increase of firing rate to threshold, the quicker the movement [Michelet et al. (2015)]. This implied high vs. low spike-count correlation when aligning spike trains with respect to movement. In contrast, we observed a correlation between dACC activity and behavior in terms of deviation from prototypical activity patterns, while we did not observe a robust link between large vs. small number of spikes emitted during 1^{st}-reward-triggered discharges, and different behaviors. This result can be well understood when considering that dACC can signal a given behavioral strategy when its activity lies close to a given prototypical state. Hence, this interpretation can be consistent with reports of increased spike count variability (and hence, of absence of defined state of activity) in dACC during periods of behavioral uncertainty [Karlsson et al. (2012)]. It can also be related to finding about a sudden reorganization of dACC activity in a new "rule encoding network state" when animals switch to a new rule [Durstewitz et al. (2010)]. Within this framework, 1^{st} reward feedback triggers specific dACC activity patterns [Balaguer-Ballester et al. (2011)] that shape the response of downstream areas such that the appropriate decision (here, switching to repetition) is taken. Deviation from these "prototypical patterns" would lead to a slower behavioral response. In addition, if the deviation of dACC discharges from their usual pattern were triggered by increased uncertainty or difficulty,

slowing the behavioral response may prevent incorrect choices (as suggested by the similar error rates between trials with fast vs. slow responses).

Interestingly, these results also suggest that the information transmitted to downstream areas cannot be mapped onto an intensity value (i.e. a single dimension), such as the magnitude of the required cognitive control, as in the case of the integrator model. Rather, the deviation from a prototypical pattern, which relates to behavioral modulation, appeared to occur in many different ways (through either an increase or a decrease of spike count, or through spike timing deviations within the heterogeneous temporal patterns of dACC neurons). This hints to the transmission of a high-dimensional representation by dACC, possibly linked to the embedding of the cognitive control signal into a specific context, or behavioral strategy [Quilodran et al. (2008); Shenhav et al. (2013); Ullsperger et al. (2014)]. Furthermore, these results also suggest a non-linear behavior for the downstream decoder. Taken together, our observations could therefore be consistent with a recent study reporting evidence for a high-dimensional, non-linear processing in lateral prefrontal cortex (lPFC, [Rigotti et al. (2013)]), an area which is likely to process dACC signals [Procyk and Goldman-Rakic (2006); Rothé et al. (2011); Shenhav et al. (2013)]. One limitation of our study is that we only characterized the dimensionality of the representation transmitted by dACC through the large differentiation, at the population level, between measures based on firing rate and measures based on (absolute) deviation from prototype. A full evaluation of this dimensionality will need future studies to evaluate the space of neuronal variability and its relation to behavioral variability in each single neuron.

Importantly, beyond the deviations of dACC spike trains from prototypical spike count, our findings indicate that deviations from prototypical temporal patterns were predictive of the monkeys' upcoming response time. This was consistent and significant in both monkeys. Furthermore, compared to the prediction based on spike count deviations, the prediction power of adapted temporal sensitivity was either equivalent (monkey P) or significantly stronger (for monkey M, which showed the most reliable relation between neural activity and behavior). This strongly suggests that the temporal patterning of single unit activity is not an epiphenomenon irrelevant to downstream network dynamics.

We note that dACC differs from other decision-making related areas such as middle temporal (MT) or orbitofrontal cortex (OFC) regarding the nature of

the relation between neuronal variability and future response time variability. Indeed, in MT and OFC, the firing rate of specific neuronal populations predicts behavioral modulation [Britten et al. (1996); Kepecs et al. (2008)]. In addition, evidence suggests that neurons in MT are decoded through integration, a process that could be reflected in LIP (lateral intraparietal cortex) activity [Huk and Shadlen (2005)], and which appears to have one-dimensional dynamics ([Ganguli et al. (2008)], see also [Latimer et al. (2015)]).

5.4 Outlook

Altogether, our results appear hard to reconcile with the hypothesis of a decoding of post-feedback dACC activity by a neural integrator. Other types of decoders could be compatible with both an increase in information through spatiotemporal coincidences and a correlation of deviation from prototypical temporal patterns to behavior. For instance, as we illustrate in Figure 2.1 c, a recurrently connected neuronal population, which maintains memory through a high-activity state, can be modulated by the temporal structure of its input [Dipoppa and Gutkin (2013b)]. Alternatively, a downstream network maintaining a memory through repetitions of sequential activations of NMDA-connected neurons, would also be sensitive to spatiotemporal patterns [Szatmáry and Izhikevich (2010)]. Our findings therefore call for a better understanding of how models of short-term memory and decision-making could reliably be modulated by a temporal input at the timescale of hundred of ms.

Also, beyond the necessity to further verify our conclusions in new data sets, future research should better investigate the cognitive factors that modulate dACC discharges. This will require a careful design of new experiments where these factors can be distinguished and measured. Indeed, the current study reports a correlation between dACC activity and the response time, but it does not give much insight about whether and how the response time modulation may favor an efficient behavioral adaptation process. One of the possible explanation for our results could for instance be a relation between dACC discharges and the motivation of the monkey. Alternatively (or in addition to the previous point), they may indicate a relation between dACC discharges and the confidence of the monkey in the appropriateness of the chosen target. Yet,

another possibility could be that dACC discharges directly reflect the attention of the monkey to the task stimuli that permit the implementation of the appropriate behavioral strategy. Note that the attentional effect can differ from a confidence effect, as one can be confidently wrong. Hence, an attentional effect would occur before the final choice is made, while a confidence effect would rather be post-decisional. In the context of our task, analyses of activation latencies, and of the strength of target-choice-related activity, suggest that dACC modulates lPFC, which in turn implements the decision about which target to touch [Procyk et al. (2000); Rothé et al. (2011); Khamassi et al. (2014)]. This would be more consistent with a pre-decisional involvement of dACC. However, another study reported the presence of post-decisional correlates in dACC [Blanchard and Hayden (2014)]. Hence, this issue will need to be investigated further in the future.

Relatedly, it is also currently difficult to determine whether dACC feedback discharges are signaling the new adapted behavioral strategy to downstream decision-and-memory areas (hence specifying this behavioral strategy [Shenhav et al. (2013)]), or whether these discharges reflect the monitoring of the extent to which a particular behavioral strategy is specified. In the first case, the modulation of the behavioral response time would be entirely dependent on the reaction of downstream areas to dACC discharges, whereas in the second case, dACC might modulate directly the speed of the decision and of the behavioral response, potentially to avoid mistakes.

Finally, it will be a difficult but extremely important goal to design a test of the causal impact of spatiotemporal structure of dACC activity on behavior. This would require to stimulate in a spatiotemporally precise fashion populations of neurons in behaving animals. While optogenetics might be a promising technique, for now it cannot be used to impose a neuron-specific temporal stimulation. This is problematic, because different neurons with the same genetic marker can show different firing patterns [Kvitsiani et al. (2013)]. Also, a simple optogenetic tagging of all strongly activated neurons (similar to techniques used in the hippocampus, for instance [Liu et al. (2012)]) would not work in our case. Indeed, different dACC populations are transiently active during a behavioral task. Hence, a satisfying and successful design of a causal experiment for investigating the function of spatiotemporal patterns of activity remains a technical challenge today.

Part III

Advances for a theoretical investigation of the function of temporal dynamics in recurrent networks

Chapter 6

Preamble: from spike train data analysis to the development of mean field methods

---

In the first part of this dissertation, we described how we tested the plausibility of different network architectures that could process dACC activity, by probing the informative features of spike trains (Figure 2.1). More precisely, we tested different types of decoding networks for feedback-related discharges, which seem to transmit information related to the appropriate behavioral strategy to be implemented in the near future. The analysis provided considerable evidence against the decoding of dACC feedback-related discharges by a simple integrator network, which would maintain the memory of dACC stimulation through a slow enough decay. However, several alternatives [Mongillo et al. (2008); Martínez-García et al. (2011); Mongillo et al. (2012); Dipoppa and Gutkin (2013b); Szatmáry and Izhikevich (2010)] may be consistent with the necessity to hold the received signals in memory, as well as with our observations:

1. The presence of a prototypical temporal pattern in dACC discharges that informs about the appropriate future behavioral strategy.

2. A slowing down for the future behavioral response when the spike trains deviate from the prototypical discharge (apparently, by either increasing or decreasing the spike count relative to the prototypical value, and/or by changing spike times relative to the prototypical spike train).

Further investigation of how a spatiotemporal decoder may make use of the information in dACC feedback-related discharges required to make some assumptions on the global structure of the decoding network. Further, it was desirable to analyze and understand the fundamental consequences of these

assumptions. Indeed, this would lead to predictions that could be cross-validated in the data.

6.1 Neuronal architectures that could plausibly support dACC activity decoding

Hence, we first needed to choose a plausible neuronal architecture, and to investigate how it could make use of the spike timing information in its input. For this, we took advantage of existing data concerning an area which could plausibly process dACC signals: the lateral prefrontal cortex (lPFC, [Shenhav et al. (2013)]). Indeed, the Local Field Potentials (LFP) recorded simultaneously in dACC and lPFC were analyzed in the same monkeys and the same behavioral task as those of our article. This analysis revealed that during the feedbacks leading to behavioral adaptation (errors or 1^{st} reward), there was a high-gamma power increase that occurred earlier in dACC compared to lPFC [Rothé et al. (2011)]. In addition, high gamma power correlations were found between the two areas during post-feedback epochs, with dACC leading dlPFC by 100-200 ms during the search period (for the 60-100Hz band, [Rothé et al. (2011)]).

Furthermore, recordings of neuronal activity in lPFC while monkeys performed the same task as for our analysis also revealed the presence of activity specific to the chosen target [Procyk and Goldman-Rakic (2006); Khamassi et al. (2014)]. More precisely, Procyk and colleagues reported the presence of choice-specific sustained activity during the delay period of the task. Hence, there were neurons whose firing rate increased more after the feedback if their "preferred" target was being chosen, and whose activity stayed elevated until the monkey made a saccade towards the chosen target. These neurons hence appeared to reflect the decision of the monkey and the memory maintenance of this decision. From a theoretical point of view, such a sustained activity could be compatible with a multistable attractor network able to maintain constant sustained activity through recurrent connections. Indeed, such a network has proven to be sensitive to its input's temporal structure [Dipoppa and Gutkin (2013b)]. Further, recent data analysis studies have shown observations compatible with (approximate) attractor networks at the level of both neuronal dynamics and correlations between neuronal and behavioral variability in frontal cortex [Balaguer-Ballester et al. (2011); Rigotti et al. (2013); Wimmer et al.

(2014)]. Hence, Balaguer-Ballester et al. (2011) showed that the trajectory followed by neuronal activity (in a neuronal space) often becomes slower close to relevant behavioral events (and therefore, when information probably had to be transmitted from and to different neuronal populations). Rigotti et al. (2013) showed that some important characteristics of multistable attractor networks designed to be able to implement complex cognitive tasks, such as a non-linear combination of the response to different cues (leading to "mixed selectivity", [Rigotti et al. (2010)]), are associated with a good performance of the animals during a memory task. Finally, Wimmer et al. (2014) showed in a saccadic memory task that correlations between neurons with sustained activity during the delay were compatible with a ring-like attractor network. In addition, Wimmer et al. (2014) found that the fine variability in the stimulus feature decoded from persistent activity at the end of the delay correlates with the memory of the animals. This strongly suggests a relation between delay activity and memory.

However, while the sustained activity during the delay has long been speculated to be influential for this memory and decision-making function [Fuster (1973)], there is controversy regarding whether there is really a causal (rather than purely correlational) link between the two [Martínez-García et al. (2011)]. Indeed, only $\approx$ 40% of neurons show sustained activity [Procyk and Goldman-Rakic (2006)], and among those only 65% are spatially tuned. Further, the firing rate of these neurons may increase or decrease over time during the delay. An alternative model proposes that the memory would rely in the loading of presynaptic calcium buffer, therefore leading to a short-term ($\approx 1s$) potentiation of synapses, and allowing memory maintenance with or without the presence of sustained activity [Mongillo et al. (2008, 2012)]. Yet, other possible models can rely on a feedforwardly activated chain of network states (e.g. [Goldman (2009)]). One implementation of such a feedforward chain in a spiking neuron network relied on repetitions of the sequential transient activations of NMDA-connected neurons (creating so-called "polychronous patterns", [Szatmáry and Izhikevich (2010)]).

Below, we review the experimental evidence which may be used to try to determine whether one model may represent the data more accurately.

6.2 Experimental evidence suggesting a causal relation between delay activity and short-term memory

Current evidence based on extracellular recordings in the frontal cortex of animals instructed to hold some items in memory often appears to be qualitatively compatible with any of the (non-necessarily exclusive...) models [Szatmáry and Izhikevich (2010); Martínez-García et al. (2011); Rigotti et al. (2013)]. Indeed, all models could be compatible with the presence of neurons which, specifically when a given set of circumstances have to be remembered, increase their firing rate during the delay and may show a temporally structured sustained discharge. However, the current implementation of the model based on polychronous patterns does not seem to yield sustained increased firing rate with an intensity that is comparable to the data (see [Szatmáry and Izhikevich (2010)]; the difficulty could be the occurence of too many patterns by chance for large firing rates). In addition, the decoding time-scale permitting robust decoding in our data appears larger than the precision at which "polychronous patterns" were activated in the simulation (which was of a few ms, a time-scale constrained by long-term spike-timing dependent plasticity). More importantly, the robustness of the memory (for longer than 2-3 s) in this "polychronous patterns" simulation required a reactivation of the pattern. However, the saturation of the transmissible information in our dACC data (see [Figure 4.2] for instance) seems to indicate that such a robust reactivation is unlikely to occur after 1s post-feedback. Even though is is unclear to what extent the above-mentioned weaknesses of the "polychronous patterns" hypothesis are specific to the published implementation, or intrinsic to the concept of the model, we feel that in the current state of knowledge this model appears rather less plausible than the others. Indeed, these other models have proven to be able to hold robust memory [Martínez-García et al. (2011); Mongillo et al. (2012)]. For the model relying on a synaptic calcium buffer only, we note that the robustness is directly determined by the short-term plasticity time-scale, which may extend until minutes [Zucker and Regehr (2002); Tsodyks and Wu (2013)]. Finally, the models that do not rely on a precise sequence of single neuron activations can probably accommodate a less constrained range of time scales for the spike timing sensitivity (e.g. see [Dipoppa and Gutkin (2013b)]). Indeed, in this case,

the spike-timing sensitivity can emerge from the non-linearity of the sustained population response, which can be shaped by many single-neuron and network connectivity characteristics.

We now turn to discussing experiments that may help determining whether a sustained firing during the delay could support short-term memory.

Three very recent optogenetic manipulation experiments actually hint towards a putative importance of such neuronal firing specifically during the delay period [Rossi et al. (2012); Gilmartin et al. (2013); Liu et al. (2014)]. This therefore appears to argue against a purely short-term facilitation-based theory, which would a priori predict that a sustained firing of excitatory neurons during the delay period would be unnecessary for successful memory maintenance.

1. Gilmartin et al. (2013) investigated the issue during trace fear conditioning which requires to hold a memory of the punishment–predictive conditioned stimulus (CS, here, a sound) during a delay (20 seconds) before a punishment is given. They used a light-activated inhibitory channel which caused an inhibition of a majority of neurons in an area (the prelimbic medial prefrontal cortex) where sustained firing had been shown during the delay between CS and punishment. While the rats underwent conditioning, using this type of inhibition specifically during the delay period – but not during the CS period – seemed to prevent the association between CS and punishment. This deficit was equivalent to the impairment observed when inhibiting the area during the whole trial (from CS to the end of the punishment period). Hence, these results are globally more consistent with the necessity of sustained firing during the delay in order to learn the association CS–US, which requires (among other things) the short-term memory maintenance of the CS.

2. Rossi et al. (2012) used a task which more specifically involves short-term memory, where mice had to press one lever (amongst two), wait while remembering which lever they had pressed, and then press the other lever to get a reward. After task acquisition, an inactivation of pyramidal medial prefrontal neurons during the waiting period (through the activation of PV interneurons), impaired task performance. Note that the medial prefrontal cortex (mPFC) contains neurons with delay-related firing [Rossi et al. (2012)].

3. Liu et al. (2014) performed similar experiments to Rossi et al. (2012) in a

similar delayed non-match to sample task. They found a similar effect of behavioral impairment when activating GABA-ergic neurons in the mPFC of mice during the delay. In addition, during the delay, they specifically inhibited pyramidal neurons with an inhibitory light-activated channel and found a behavioral impairment.

These studies emphasized the role of any type of item-specific frontal firing during the delay for short-term memory (including complex trajectories implying both an increase and a decrease of firing rate). However, these studies do not specifically argue for a function of sustained, stable firing rates during the delay. To the best of our knowledge, a study which would specifically manipulate the activity of the neurons which preferentially fire while a given item is being remembered (similar to what was done for context-specific activity in hippocampus [Liu et al. (2012)]) is still missing. However, there exist again an imperfect and indirect evidence for some relevance of an increased firing rate in item-specific neurons for the memory of this item. Indeed, the neurons in one hemisphere are more likely to have their "preferred item" (i.e. the item leading to higher sustained activity during memory) contralaterally. Remarkably, unilateral lesions [Funahashi et al. (1993), in monkey lPFC] or inactivations [Hanks et al. (2015), in rat frontal cortex] lead to a contralateral deficit, i.e. a bias of the memory and/or the decision towards the ipsilateral item wich is preferred by the neurons of the other hemisphere. This appears consistent with an encoding of the memory for one item by a larger firing rate in the delay activity of one subpopulation of lPFC neurons, which would compete with other subpopulations whose sustained firing encodes the memory of other items. In contrast, such findings are harder to explain when assuming that the memory is encoded in lPFC through a complex firing rate trajectory involving both an increase of firing rate at some times, and a decrease of firing rate at other times.

6.3 A hypothesis for the decoder of dACC that is compatible with the current literature

Based on this (incomplete) evidence, we therefore decided to explore further the assumption that lPFC was indeed maintaining the memory of the decision through sustained firing, until the animal can express its choice by making a

saccade. In addition, we assumed that lPFC participates in making the decision about which target to touch next after the monkey receives a feedback. More specifically, we reasoned that this could occur through competitions between different pools of neurons which code for different decisions.

Hence, the network would be composed of four populations of recurrently connected neurons, with inhibitory connections between these populations. Each of these populations would possess two stable states of sustained activity during which either the low or the large firing rate would be maintained through recurrent connections. This architecture could therefore produce firing rate profiles that would be compatible with the observations in lPFC [Brunel and Wang (2001); Martínez-García et al. (2011); Dipoppa and Gutkin (2013b)]. As we mentioned previously, this type of attractor network can be sensitive to the temporal structure of its input [Dipoppa and Gutkin (2013b)]. Also, the state of sustained activity of such a network may be destabilized if the neuronal population receives an input that is too strong, which may help explaining our observations of increased behavioral response time if dACC spike trains have either too many or too little spikes. For instance, sustained activity in a bistable network can be destabilized if the external input synchronizes all the neurons of the active population [Gutkin et al. (2001); Dipoppa and Gutkin (2013a)]. We note that given the non-linearity of the dynamics of such networks, several mechanisms could explain this phenomenon (e.g. relying on shunting through increased conductance, or on rebound inhibition [Gutkin et al. (2001)).

We hypothesized a role for dACC in sending a signal specifying the behavioral strategy, which in our case is equivalent to a signal specifying whether to avoid or to touch again one of the targets that was chosen in the past. However, the identities of the previously touched targets were rather weakly encoded in the firing rates of lPFC neurons, which delay activity is largely related to the chosen target to which the monkey will saccade in the future [Procyk and Goldman-Rakic (2006)]. Previously touched targets also appeared to not be very strongly encoded in dACC signals, which rather reflect the need for cognitive control and different internal states corresponding to different behavioral strategies [Procyk et al. (2000); Shenhav et al. (2013)].

Still, we reasoned that the adaptation properties of single lPFC neurons may create temporal patterns of activity, or "hidden" excitability states that are not obviously apparent in the spiking activity. These adaptation states may identify whether the neurons have been activated in the past. Note that adaptation was

indeed shown (through simulations) to be compatible with bistable dynamics for a recurrent neuronal population [Theodoni et al. (2011)].

Hence, a "hidden" memory of which targets were touched, and of when they were touched, may actually be present in a network of adapting neurons in which a sustained firing rate indicates the future touched target. Note that conceptually similar ideas had been suggested in the past and studied through simulations of recurrent networks with short-term plasticity [Buonomano and Merzenich (1995)].

The adaptation properties of the neurons can occur at the level of both the membrane properties (on time scales extending from milliseconds to 20 seconds [La Camera et al. (2006); Lundstrom et al. (2008); Pozzorini et al. (2013)]), and at the level of synaptic release probability (generally on time scales extending up to ≈ 1 second [Mongillo et al. (2008); Wang et al. (2006)]). Interestingly, concerning the time scales of membrane properties adaptation, the power law decrease of the amplitudes of the different time-scales of adaptation [Pozzorini et al. (2013)] could be compatible with a decay of the performance with the duration of the memory. This is a hallmark of working memory [Liu et al. (2014)]. Also, making this hypothesis of a memory of previously touched targets within the adaptation state of the choice-related neuronal population may be compatible with a function of the temporal structure of the behavioral-strategy input received from dACC. Indeed, it appears conceivable that a given temporal input could be particularly well suited to excite a population undergoing a specific adaptation state. This adaptation state could be specific to a given delay since the population had been last activated (and thus to a given delay since the associated target had been last chosen).

For instance, the prototypical 1^{st} reward temporal pattern emitted by some dACC neurons might be well–suited to excite a population of neurons that had been activated relatively recently in the past ($\approx 1.5s$ ago, which is the delay between saccade and reward), and less well–suited to activate a population of neurons that had been activated much before (during previous trials, more than $4s$ ago). The hypothesized functioning of such a circuit is depicted in Figure 6.1. Note that during repetition, there was no correlations between dACC post-feedback gamma power and lPFC post-feedback gamma power [Rothé et al. (2011)], which may indicate that another implementation mechanism takes over during this period where the required cognitive control is low.

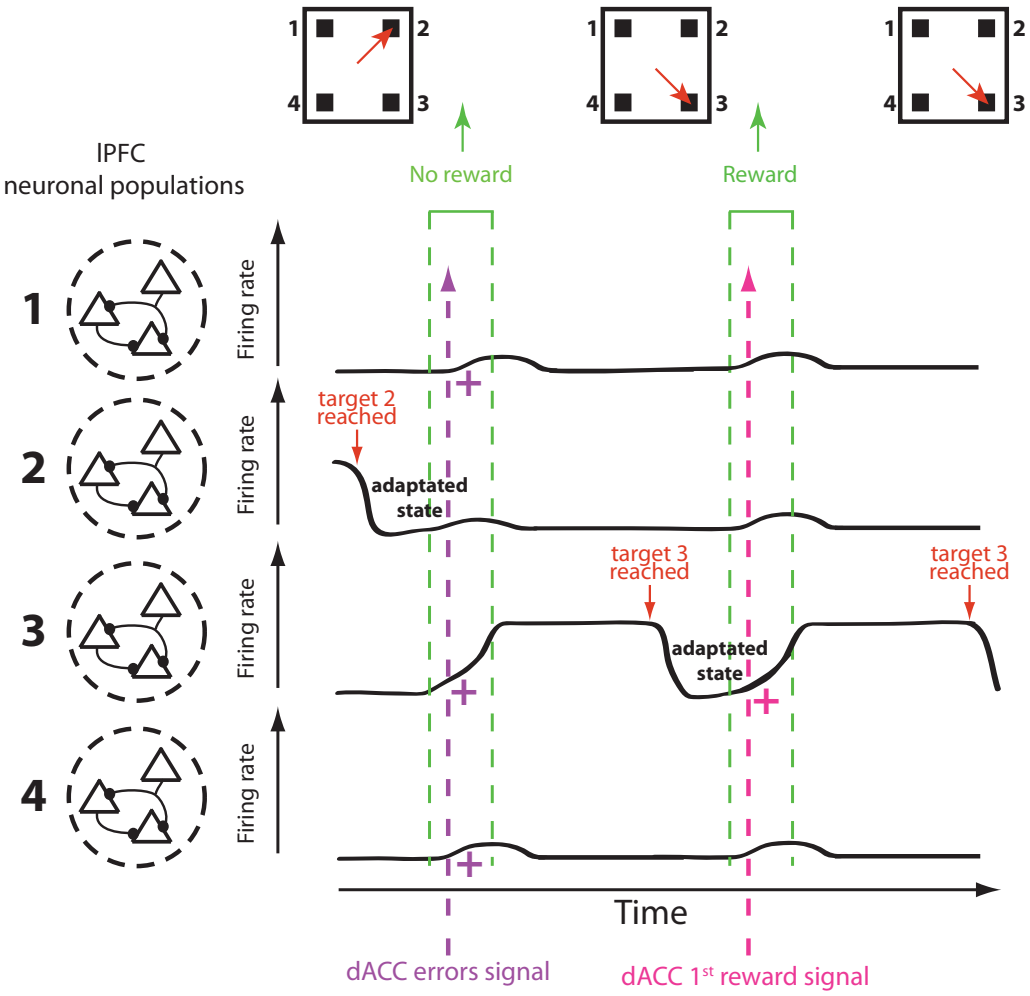


Figure 6.1: *A hypothesis for the functioning of lPFC and its modulation by dACC during the problem solving task. Top*: Schematic showing an example of action sequence during the solving of a problem (see section 3.1), the rewarded target being on the bottom right. The monkey first touches the top right target, and is therefore not rewarded. The monkey then tries the bottom right target, selects it, and gets rewarded. Therefore, he then reselects the same target (bottom right), as the monkey knows that the same target is rewarded several times in a row. *Bottom*: Schematic illustrating a putative network architecture which would produce firing rate profiles similar to those observed in lPFC [Procyk and Goldman-Rakic (2006)]. Four groups ("populations") of neurons were observed, corresponding to the four targets. Note that these populations could be competing through mutual inhibition (which we did not illustrate for simplicity). When the monkey starts making a decision (at the beginning of a problem, or after a feedback), the firing rate begins to rise and becomes sustained and higher in the neuronal population corresponding to the chosen target, and the firing rate of this population drops once the monkey expresses its choice by making a saccade to the corresponding target. This neuronal population is therefore in an adapted state when the corresponding feedback is given. We hypothesize that the error signal provided by dACC might be better suited to excite neuronal population that are not in an adapted state, while the 1^{st} reward signal might be better-suited to excite the population that is in the more adapted state.

6.4 How to study the hypothesized network for dACC decoding?

It is a scientific question by itself to determine whether and how a network such as the one sketched in Figure 6.1 can be built, and another question to determine whether it is in agreement with the data beyond the qualitative arguments detailed above, that led us to imagine such a circuit (which is, of course, only one possibility among others).

We have been trying to work on the first of these two questions. Understanding how temporal structure in the input dynamically modulates the response of (non-linear) recurrent neuronal networks with adapting neurons is actually still challenging from a theoretical point of view, as we will discuss and review in the next chapters.

Before this, we would like to elaborate a little bit on the fact that, as often in research, the real challenge was going well beyond technical difficulties and was mostly about determining the right approach and orientation to advance in the resolution of the problem. Indeed, it would have been possible to take, since the beginning, a simulation-oriented approach. Simulations relating to similar problems have indeed been successfully implemented in the past [Buonomano and Merzenich (1995); Dipoppa and Gutkin (2013b)], and they provide a necessary proof of principle. However, one might feel that the understanding of the mechanism at stake stays obscured by the complexity of the simulated system [Dipoppa and Gutkin (2013b)]. In addition, it can become difficult to compare the simulation and data (beyond the features that the simulation was built for reproducing), as it can be hard to isolate a strong (i.e., probably robust to the relaxation of the simplifications made in the model) constraint of the modeled mechanism that could be taken as a prediction. For these reasons, and also probably because the relative appeal of pure simulations vs. analytics is a matter of one's personal way to reach a satisfying feeling of intuitive understanding, the analytical approach was pursued.

However, there are also numerous pitfalls when trying to phrase the problem in an analytically tractable form. Indeed, even when starting with the minimal model required to approximately reproduce neuronal dynamics under relatively mild assumptions, the complexity of the equations can prevent any intuitive analysis of the mechanisms at stake. And indeed, we initially stated the problem

in such a complex way. We attempted to exactly account for temporal correlations beyond the trial-averaged firing rate. Several self-consistent nested integral equations were written. They turned out to be much more complicated to solve numerically than the simulation that they were describing, and they were not (in our view) bringing any intuition for how the network's dynamics was arising. This approach was therefore abandoned, because it seemed to be irrelevant to our initial question. Rather, such an approach is relevant to the question of the probabilistic mathematical reformulation of a given network model.

We then considered the other extreme: taking an arbitrary phenomenological firing rate model, and investigating its dynamics and its computation abilities. While this led to valuable insights on the computational potential of the particular model chosen, we found it difficult again to isolate a good prediction of the model. Indeed, it was unclear how to determine whether and how the results depended on the simplifications of the model.

We finally decided to use a simplified and approximate mean-field model, while retaining an analytical derivation of the model's dynamics. This allowed us to clarify the assumptions made in order to derive the final formula from the equations for a single-neuron model that can be fitted to recorded neurons [Pozzorini et al. (2013)]. Hence, this approach permits to have a good idea of how the simplifications made affect the results. Ultimately, such a method should be amenable to providing an intuitive comprehension of how an external temporal input may interact with the internal adaptation properties of a recurrently connected bistable population, to favor its switch to a high-activity state (Figure 6.1).

The following chapters describe how we derived and tested this approximate mean-field method.

Chapter 7

Introduction: how to analyze the dynamical response of recurrent adapting networks of neurons?

Over the last decades, the study of the dynamics of coupled population of neurons has attracted a lot of attention from the scientific community as both a theoretical challenge [Sompolinsky et al. (1988); Abbott and van Vreeswijk C (1993); Van Vreeswijk et al. (1994); van Vreeswijk and Sompolinsky (1996, 1998); Brunel (2000); Gerstner (2000); Renart et al. (2007, 2010); Mongillo et al. (2012); Sussillo and Barak (2013); Wainrib and Touboul (2013)], and as a successful tool to approach the question of the generation of internal representations and behavioral outputs by the brain [Seung (1996); Seung et al. (2000); Compte (2000); Brunel and Wang (2001); Wong and Wang (2006); Balaguer-Ballester et al. (2011); Rigotti et al. (2013); Haefner et al. (2013); Wimmer et al. (2014, 2015)]. More generally, theoretical studies have been extremely insightful for integrative neuroscience, as qualitative reasoning or approaches purely based on simulations soon lead us to face the difficulty of grasping how a global behavior can emerge from a complex set of many interacting elements. Hence, many theoretical approaches reduced the complexity by self-consistently computing some moments, or even the whole distribution, of relevant variables among a population of neurons that are similar in their dynamical properties and their connectivities. These approaches typically require a simple enough model for the single neuron, such as a binary units [van Vreeswijk and Sompolinsky (1996); Renart et al. (2010)] or integrate-and-fire models [Brunel (2000); Renart et al. (2007)]. These models do share important features with real neuronal networks, with spike-based interactions between neurons and integration-like dynamics, and can allow a quantitative match to the steady-state firing rate [Arsiero et al. (2007)]. Hence, successful and discerning comparisons between neuronal population analyses for

these types of models, and data, could typically be made for correlations between neurons [van Vreeswijk and Sompolinsky (1996); Renart et al. (2010)], for fast and strong interplay between excitatory and inhibitory populations [Brunel (2000); Brunel and Wang (2003)], and for the characteristics of the stationary (or quasi-stationary) response patterns [Compte (2000); Brunel and Wang (2001); Renart et al. (2007); Mongillo et al. (2012); Wimmer et al. (2014)].

However, the classical single-neuron models amenable to mean-field analysis do not in general allow a quantitative match of precise spike times when fitted against recorded pyramidal neurons receiving complex non-stationary input current. Indeed, successful fitting of the time-dependent response of pyramidal neurons to non-stationary synaptic-like input at the soma often requires to account for neuronal adaptation on multiple time scales [La Camera et al. (2006); Lundstrom et al. (2008); Kobayashi et al. (2009); Pozzorini et al. (2013)], with effects that cumulate over spikes and that cannot typically be considered as stationary. This adaptation incorporates the effect of both hyperpolarizing currents triggered by spikes, and increases of the voltage threshold at which spikes are being generated [Pozzorini et al. (2013)]. These characteristics significantly complicate the derivation of population-wide statistics [Gerstner (2000); Gerstner and Kistler (2002); Gerstner et al. (2014)]. Indeed, the mathematical treatment generally requires to approximate the adaptation either by considering a dependence on the last spike time only (so-called renewal theory [Wilson and Cowan (1972); Gerstner (2000); Toyoizumi et al. (2009)]), or by averaging adaptation variables while assuming that they are slow relative to the neuronal dynamics (interspike interval or membrane voltage dynamics [La Camera et al. (2004); Gigante et al. (2007); Muller et al. (2007); Farkhooi et al. (2011); Hertäg et al. (2014)]). Given that in pyramidal neurons, the amplitude of adaptation effects appear to follow a power law on time-scales ranging from milliseconds to seconds [Lundstrom et al. (2008); Pozzorini et al. (2013)], these assumptions are expected to be violated for these excitatory neurons.

Recently, a new mean-field approach was developed based on a non-linear single neuron model with stochastic threshold that can be fitted to a time-varying input current. This model, which belongs to the class of Generalized Linear Models (GLM) for single neurons, incorporates both voltage and threshold adaptation on multiple time scales [Pozzorini et al. (2013)]. The mathematical analysis allowed to compute the average response of a neuron to different repetitions of a non-

stationary stimulating current with frozen noise [Naud and Gerstner (2012a)]. In addition, recent developments allowed to compute the firing rate of a connected populations of a finite number of neurons, where all neurons receive the same current [Deger et al. (2014)]. Hence, in this framework, it is not possible to analyze the impact of the within-population, between-neuron differences in the received fluctuating synaptic input.

However, there is evidence that the firing of neuronal populations in the neocortex is significantly driven by the presence of fluctuations in the synaptic current that are mostly unshared from one neuron to the next (leading to irregular and nearly uncorrelated neuronal firing [van Vreeswijk and Sompolinsky (1996); Shadlen and Newsome (1998); Holmgren et al. (2003); Rudolph et al. (2007); Renart et al. (2010)]). Further, several studies suggest that changes in the amplitude of these fluctuations could be relevant for driving the response of biological neuronal networks in behaving animals. For instance, modeling studies suggested that the increased irregularity of interspike intervals during sustained activity in frontal cortex (i.e. while the animal holds an item in memory, compared to baseline [Compte et al. (2003)]) could be explained if the sustained activity was caused by an increased amplitude of the current fluctuations [Renart et al. (2007); Mongillo et al. (2012)]. Indeed, if instead the increased activity would be caused by an increase in the mean current received by the neuron, an increase in the regularity of the discharge would be expected: after each spike, the time of the next spike would principally depend on how fast the voltage increases from reset to threshold [Schwalger and Lindner (2013); Gerstner et al. (2014)]. Furthermore, pyramidal neurons in prefrontal cortex were found to reliably respond to changes in the variability of their input current [Arsiero et al. (2007)]. Finally, recently, the relevance of fluctuation-driven dynamics was further strengthened by a recent theoretical study [Lim and Goldman (2013)], where such dynamics were proposed as a robust mechanism which could plausibly permit to implement an approximate integration of the synaptic input received by a recurrent network.

Treating analytically these dynamical changes in the amplitude of the fluctuations is still rather challenging, even for networks of simple neurons without adaptation. Hence, many studies assume that the level of fluctuations does not vary over time [Brunel (2000); Gerstner (2000); Ostojic and Brunel (2011)], or only focus on how a discrete change of the level of fluctuations impacts the steady-state response [Renart et al. (2007); Mongillo et al. (2012)].

In addition, for different types of integrate-and-fire neurons with reset but without adaptation, some analytical formulas for the time-dependent response to changes in the amplitude of fluctuations do exist, but they appear to be restricted to a linear, or weakly non-linear, response in the presence of white noise [Amit and Brunel (1997); Brunel and Hakim (1999); Lindner and Schimansky-Geier (2001); Fourcaud-Trocmé and Brunel (2005); Tetzlaff et al. (2012); Helias et al. (2013); Kriener et al. (2013)]. Also, a phenomenological model was recently derived to handle the non-linear response to both mean and fluctuation-driven inputs analytically [Tchumatchenko and Wolf (2011)]. However this model had no reset and no spike frequency adaptation, suggesting that it may have a limited explanatory power of the dynamical response of pyramidal neurons to non-stationary input [La Camera et al. (2006); Lundstrom et al. (2008); Kobayashi et al. (2009); Pozzorini et al. (2013)].

Extending the above-mentioned approaches to adapting neurons with dynamical modulation of the amplitude of the fluctuations is rather challenging, because the diverse adaptation variables indirectly follow these fluctuations at different time scales [Hertäg et al. (2014)]. In addition, in a recurrent network, adaptation introduces temporal correlations in the input current which are also hard to treat, but that can have large effects on the dynamical response of neurons ([Brunel et al. (2001); Fourcaud-Trocmé et al. (2003); Brunel and Latham (2003); Köndgen et al. (2008); Moreno-Bote and Parga (2010); Tchumatchenko and Wolf (2011)], but see [Alijani and Richardson (2011)]).

Hence, there are several technical difficulties for deriving analytical formulas accounting for adaptation within a mean-field analysis of a recurrent spiking population undergoing changes of both the neuron-averaged input, and the neuron-independent variability of the input. In addition, beyond deriving mathematical expressions, the aim of the analysis should be to bring an intuition on how the single neuron properties can shape the network's response and play a role in brain processing. This requirement for an explanatory power of the analysis calls for the use of clever approximations, that would considerably reduce the complexity of the formulas while preserving important features of the neuronal response.

Here, we tackle these issues using a generalized integrate and fire single neuron model with adaptation, that can capture the dynamical response of cortical neurons [Mensi et al. (2012); Pozzorini et al. (2013)]. We propose an approach which takes root on an approximate expression derived for the average

firing rate of this neuron in response to repetitions of a non-stationary stimulating current with frozen noise [Naud and Gerstner (2012a)]. This expression accounts for the adaptation effects that can be estimated through the history of past average firing rate. We extend this formula by making an average over the different inputs received by different neurons, while still making the assumption that time-dependent firing rates can account for spatiotemporal correlations [Brunel (2000)]. We show why, in most cases, the distribution of input values (and therefore, the distribution of subthreshold voltage values) over the different neurons can be taken as Gaussian [Destexhe et al. (2003)]. We derive how the parameters of this Gaussian vary over time as the input evolves, and we make use of this result to compute the average non-linear neuronal response. Our approximations are valid when neurons fire asynchronously and irregularly, as often observed in the neocortex [Shadlen and Newsome (1998); Compte et al. (2003); Renart et al. (2010)]. Our analysis also takes into consideration the correlations between the fluctuations of the membrane potential due to synaptic input, and the fluctuations of the adaptation variables, for each single neuron. To do this, we linearize the adaptation variables (after averaging over the different responses reached for different repetitions of a deterministic stimulation). We stress that we still treat the spiking non-linearity analytically, hence preserving many relevant non-linear features of the population response. At the end, we reach rather simple mathematical expressions that can be written in the form of non-linear differential equations. Furthermore, the formulas for the steady-state response can be written as simple coupled transcendental equations. Finally, for a single recurrent population, the steady-state response boils down to the Lambert-W function, which has well-defined solutions.

Chapter 8

Derivation of approximate expressions for the dynamics of recurrent adapting networks of neurons

---

In this chapter, we will start by describing the model of single neuron dynamics that we use, and by explaining to which extent and in which conditions it is found to be an accurate description for the dynamical response of cortical neurons (in section 8.1). We then explain how to derive an approximate analytical formula for the expected activity among a subpopulation of neurons with similar parameters, in a regime where they fire asynchronously and irregularly (in section 8.2). Finally, we explain the characteristics of the network that we used to compare the analytical formulas to simulations (in section 8.3).

8.1 Single neuron model

We used a model belonging to the class of "Generalized Linear Model" (GLM), in which a (filtered) input and a filtered spiking history combine to define a spiking probability at each time.

8.1.1 Spiking probability of the GLM

The adapting GLM model states that for any small interval dt around the time t, the probability that the neuron model number i emits a spike is $\lambda_i(t)\ dt$, where the firing rate $\lambda_i(t)$ is defined in the following way:

$$\begin{aligned} \lambda_i(t) &= \lambda_0 \exp\left(h_i(t) + \eta * S_i(t)\right) \\ S_i &= \sum_{\{t_i^k\} \leq t} \delta\left(t - t_i^k\right) \end{aligned} \tag{8.1}$$

Here, λ_0 is a baseline firing rate; h_i is a (filtered) driving input; δ is the dirac distribution, and $*$ is the convolution operator. Finally, $\left\{t_i^k\right\} = \{t_i^1, t_i^2, ...\}$ is the ensemble of spike times emitted by the considered neuron (number i), and η is a so-called spike history filter which accounts for refractory and adaptation effects that modulate the spiking probability depending on the spiking history [Gerstner and van Hemmen JL (1993); Gerstner (1995); Truccolo et al. (2005); Pillow et al. (2008)].

We would like to stress several important features of this model. First, the spiking mechanism has an exponential non-linearity, much alike the exponential rise of voltage close to the spiking threshold in recorded neurons [Jolivet et al. (2006); Badel et al. (2008)]. Second, the definition of the spike-history filter allows for both very strong refractoriness and adaptation. Indeed, this filter can take very negative values at short time-lags (hence effectively preventing any spiking just after a spike was emitted), and can also incorporate longer time-scales (hence leading to a modulation of the firing probability depending on the more ancient spiking history).

8.1.2 Interpretation of the filters of the GLM in a current-based approximation of the single-neuron somatic dynamics

The above-mentioned model may be used as a purely phenomenological description of spiking (as in e.g. [Pillow et al. (2008); Park et al. (2014)]), or may be matched to some biophysically defined neuronal characteristics [Mensi et al. (2011, 2012); Pozzorini et al. (2013)].

Indeed, the mathematical definition of the firing probability ($\lambda \ dt$, see Equation 8.1) can be reinterpreted as an exponential function of the distance between the somatic subthreshold voltage $V_{subthld}$, and a (dynamic) voltage threshold for firing T_{volt}. More precisely, the firing probability depends on the magnitude of this distance compared to the intrinsic noise of the neuron ΔV (in

voltage units).

Hence, for a single unit i:

$$\lambda_i(t) = \lambda_{biophys} \exp\left(\frac{V_{subthld,i}(t) - T_{volt,i}(t)}{\Delta V}\right) \tag{8.2}$$

ΔV approximately accounts for the intrinsic stochasticity of single neurons, which is due to various factors such as the finite number of channels, the stochastic nature of the opening of these channels, and the finite number of ions in a neuron Diba et al. (2004)). Because of this intrinsic noise, neurons fire slightly differently in response to different repetitions of the same current stimulus [Mensi et al. (2012); Pozzorini et al. (2013)], with a discharge that is more stochastic when the neuron is not strongly driven by the stimulus. In contrast, when the ratio $\frac{V_{subthld}-T_{volt}}{\Delta V}$ becomes close to zero or even positive, the firing probability should increase exponentially [Jolivet et al. (2006); Badel et al. (2008); Mensi et al. (2011)], leading to an almost deterministic spike emission.

In order to give a biophysical interpretation of Equation 8.1, we need to decompose the adaptation effects into changes in the voltage threshold for firing, and changes detectable at the level of the membrane potential [Mensi et al. (2012)]. Hence, adaptation effects are split between (i) a spike-triggered increase (relative to a baseline T_0) of the voltage that needs to be approached in order for the neuron to fire: $T_{volt,i}(t) = \eta_T * S_i(t) + T_0$; and (ii) a hyperpolarizing current that is generated intrinsically each time a spike is triggered: $\eta_{curr} * S_i(t)$. Note that this hyperpolarizing current implements both a reset, and adaptation effects. Then, we can rewrite the dynamic firing probability in response to an input current $I(t)$, by the mean of a membrane filter κ representing the low-pass properties of leak currents:

$$\begin{aligned} T_{volt,i} &= -\eta_T * S_i(t) + T_0 \\ V_{subthld,i} &= \kappa * (I_i(t) + \eta_{curr} * S_i(t)) \end{aligned} \tag{8.3}$$

Hence, we can equate, between Equation 8.1, Equation 8.2 and Equation 8.3:

$$\begin{aligned} h_i(t) &:= \frac{\kappa}{\Delta V} * I_i(t) \\ \eta &:= \frac{\eta_T}{\Delta V} + \frac{\kappa}{\Delta V} * \eta_{curr} \\ \lambda_0 &= \lambda_{biophys} \exp\left(\frac{-T_0}{\Delta V}\right) \end{aligned} \tag{8.4}$$

Note that the intrinsic unreliability is rather small (at least when estimated from in vitro recordings). Quantitatively, the fitting gives $\Delta V \approx 0.5-1mV$ while the range of the voltage fluctuations (i.e., the range of the fluctuations of $V_{subthld}$ or of $V_{subthld}-T_{volt}$) can be on the order of $10-20mV$ (see [Pozzorini et al. (2013)], their Fig.3b Table S1 and Fig. S6d). Hence, the dynamics are truly driven by the changes in external input and the membrane response to these changes.

The previously described biophysically interpretable model of neuronal dynamics defines the dynamics in terms of changes in current or voltage thresholds. However, the dynamics actually result from opening or closing of channels, which change the conductance of the neuron. This change of conductance indirectly leads to a change of current after multiplication of the conductance by the difference between the membrane potential and the reversal potential of the considered ion. Though in principle a change of conductance is not exactly equivalent to a change of current as the current change is independent of the change of membrane potential, current and conductance changes can be approximately related as long as the reversal potential of the ion is far away from the values of voltage reached by the membrane [Richardson and Gerstner (2005); Gerstner et al. (2014)]. We summarize here the argument, based on the negligibility of the product of two deviation terms (one for the conductance, and the other for the voltage). Let us consider two different time-dependent conductances $g_1(t)$ and $g_2(t)$, as well as a (constant) leak conductance g_L. Their respective reversal potential are E_1, E_2 and E_L. The dynamics of the voltage V at the soma reads:

$$
\begin{aligned}
C\frac{dV}{dt} &= -g_L(V-E_L) - g_1(t)(V-E_1) - g_2(t)(V-E_2) \\
&\Longleftrightarrow \\
C\frac{dV}{dt} &= -g_L(V-E_L) + \langle g_1\rangle(V-E_1) + \langle g_2\rangle(V-E_2) \\
&\quad - (g_1(t)-\langle g_1\rangle)(V-E_1) - (g_2(t)-\langle g_2\rangle)(V-E_2) \\
&\Longleftrightarrow \\
C\frac{dV}{dt} &= -(g_L+\langle g_1\rangle+\langle g_2\rangle)\left(V - \frac{g_L E_L + g_1 E_1 + g_2 E_2}{(g_L+\langle g_1\rangle+\langle g_2\rangle)}\right) \\
&\quad - (g_1(t)-\langle g_1\rangle)(V-E_1) - (g_2(t)-\langle g_2\rangle)(V-E_2) \\
&\Longleftrightarrow \\
C\frac{dV}{dt} &= -g_0(V-E_0) - (g_1(t)-\langle g_1\rangle)(V-E_1) - (g_2(t)-\langle g_2\rangle)(V-E_2)
\end{aligned}
\tag{8.5}
$$

Where the angular brackets denote averaging over time, $g_0 := g_L + \langle g_1 \rangle + \langle g_2 \rangle$ is an effective input-regime-dependent "leak" conductance, and $E_0 := \frac{g_L E_L + g_1 E_1 + g_2 E_2}{g_0}$ is an effective input-regime-dependent equilibrium potential. More specifically, if one would fix $g_1(t)$ to $\langle g_1 \rangle$ and $g_2(t)$ to $\langle g_2 \rangle$, then the voltage V would converge to E_0.

Finally one can write:

$$\begin{aligned} C\frac{dV}{dt} = &-g_0 (V - E_0) - (g_1(t) - \langle g_1 \rangle)(E_0 - E_1) - (g_2(t) - \langle g_2 \rangle)(E_0 - E_2) \\ &- (g_1(t) - \langle g_1 \rangle)(V - E_0) - (g_2(t) - \langle g_2 \rangle)(V - E_0) \end{aligned} \tag{8.6}$$

In the sum that constitutes the right hand side of Equation 8.6, the two last terms can be considered as small as long as $\forall i \in \{1, 2\}$, $(V - E_0) << (E_0 - E_i)$. This is often the case as the membrane potential often oscillates between $-40/-60$ mV (E_0 being situated in between), while many common ions such as K^+ or Na^+ have reversal potential that are very away from these voltages (≈ -77 mV for K^+ and $\approx +55$ mV for Na^+ [Gerstner et al. (2014)]).

Under this approximation, we can write:

$$C\frac{dV}{dt} \approx -g_0 (V - E_0) - (g_1(t) - \langle g_1 \rangle)(E_0 - E_1) - (g_2(t) - \langle g_2 \rangle)(E_0 - E_2) \tag{8.7}$$

In this last expression, one can see that there is no multiplication between the voltage and a time-dependent factor: in other words, the neuron behaves as if it were current-driven. Note that the neuron now possesses an effective leak current and reversal potential, which depend on the total conductance of the neuron when it is in a given input regime.

8.1.3 Validity domain of the GLM for describing single neuron's response to somatic current injections

For in-vitro current-clamp recordings of layer 5 pyramidal neurons and interneurons in the somatosensory cortex, the current-based description of neuronal dynamics of Equation 8.3 actually permitted a quantitative fit of both the subthreshold voltage $V_{subthld}$ and of the spike times (within a precision of a few ms) [Mensi et al. (2012); Pozzorini et al. (2013)]. This fit is significantly better than the fit allowed by simpler leaky integrate-and-fire models, which can

only capture the long term firing rates between 0-10Hz, without reproducing precise spike times in pyramidal neurons [Kobayashi et al. (2009)]. This occurs because pyramidal neurons possess strong adaptation properties extending from short to long time-scales. A similar result was found in pyramidal neurons of frontal cortex in vitro (i.e., Equation 8.3 could be successfully fitted to data from [Thurley et al. (2008)] when including adaptation on multiple time scales, personnal communication from C. Pozzorini).

The above-mentioned fits were realized with realistic and rich synaptic-like stimulating currents which could produce very large fast modulations of the firing rates. For instance, two spikes could occur within a few ms and then be followed by a silence period of several hundreds of ms [Mensi et al. (2012); Pozzorini et al. (2013)]. However, these stimulating currents were driving the neurons over a more limited range of long term firing rate regimes (e.g., 10-second average rates were mostly constrained between 2 and 10 Hz, see [Mensi et al. (2012)]). This does not cover the whole range of stationary firing rates that can be sustained by pyramidal neurons (which spans values from 0 to $\approx$ 25 Hz [Arsiero et al. (2007)]). Fitting the pyramidal neuron's response over their whole range of steady-state firing rates actually necessitates to enrich Equation 8.3 with a non-linear modulation of the spiking threshold by the voltage [Pozzorini et al. (2015); paper under review at PLOS computational biology]. However, it is possible to locally (over an ensemble of input regimes which drive the neurons at steady state firing rates spanning a range of about 8 Hz) remap this more complicated model on the simpler model described by Equation 8.3 [Pozzorini et al. (2015); paper under review at PLOS computational biology]. Hence, different input regimes leading to drastically different steady-state firing rates can then be separately handled by an equation of the type of Equation 8.3, each of them necessitating to use a particular set of parameters and filter shapes [Mease et al. (2014)].

As a conclusion, the model proposed in Equation 8.3 can to some extent capture the dynamical response of both pyramidal neurons and interneurons with a fixed parameter set. For pyramidal neurons, the fit is restricted to a moderate ($\approx 8Hz$) range of steady-state firing rates (while still permitting a very large range of fast firing rate modulations) .

Note that, in vivo, the precise amplitude of the adaptation effects may be modified by the presence of neuromodulators [Satake et al. (2008); Thurley et al. (2008)]. However, adaptation has still been observed during in vivo recordings (as shown in anesthetized animals [Degenetais (2002)]). Also, as argued in the

previous section, the high-conductance state of neurons in vivo [Destexhe et al. (2003)] should still be well-modeled by a current-based description such as the one proposed in Equation 8.3. This should at least work well when trying to fit the response of neurons during a given regime of synaptic bombardment. However, a study performing a quantitative fit in vivo is still lacking. Indeed, such a fit would also require to estimate the synaptic input received by the neuron under study (while, in vitro, the stimulation can be carefully controlled through the electrode in the absence of synaptic input).

8.1.4 Modeling the synaptic input and its transmission to the soma through passive dendrites

For the synaptic input, we work at the same level of approximation as for the soma model (Equation 8.3) and we therefore adopt a linear current-based description [Richardson and Gerstner (2005); Gerstner et al. (2014)]. Note that while some non-linear synapses such as those of the NMDA type do exist, it is in principle possible to linearize their response around a given synaptic input regime [Brunel and Wang (2001)]. Given the original non-linear synaptic equations, it can even be possible to compute analytically a best-approximating linear filter analytically [Thomas et al. (2000)]; hence, this approach may permit a possible extension of our framework to non-linear synapses if needed. Note that this approach might also permit some extension to synapses undergoing non-linear short-term plasticity, provided this short-term enhancement or depression dynamics could be approximated by a linear filter within some restricted regime of synaptic input [Thomas et al. (2000); Mongillo et al. (2012)].

We also assume that the dendritic processing can be approximated as a passive conduction of the current, leading to the transmission at the soma of the sum of the different (filtered) inputs. While this may be a crude approximation whose impact is currently hard to measure, we note that some studies suggest that it might be a good approximation in the high conductance state that neurons typically experience in vivo. Indeed, the conduction of synaptic input was found to be more synapse-location-independent in the high-conductance state [Destexhe et al. (2003)].

Hence, the synaptic current received at the soma $I_{syn,\,i}$ by the neuron i will

be taken as a sum over all synapses s associated with a spike train $S_{s,\,i}$ and responding to each presynaptic spike by the current time course (i.e., the impulse response) F_s:

$$I_{syn,\,i} = \sum_{s=1}^{N_{s,\,i}} \left(F_{s,\,i} * S_{s,\,i}\right)(t) \tag{8.8}$$

Hence, we can express the voltage fluctuations $h_i(t)$ (see Equation 8.1, Equation 8.3 and Equation 8.4) that are generated in neuron i by this synaptic input at the soma:

$$\begin{aligned} h_i(t) &= \frac{\kappa}{\Delta V} * I_{syn,\,i}(t) \\ &= \frac{\kappa}{\Delta V} * \left(\sum_{s=1}^{N_{s,\,i}} \left(F_{s,\,i} * S_{s,\,i}\right)(t) \right) \\ &= \sum_{s=1}^{N_{s,\,i}} \left(\frac{\kappa}{\Delta V} * F_{s,\,i} \right) * S_{s,\,i}(t) \\ &= \sum_{s=1}^{N_{s,\,i}} F_{s,\,i}^{tot} * S_{s,\,i}(t) \end{aligned} \tag{8.9}$$

where we defined a combined leak-and-synapse filter $F_{s,\,i}^{tot} := \frac{\kappa}{\Delta V} * F_{s,\,i}$.

8.2 Dynamical computation of the firing rate distribution in a recurrent network of GLM neurons

In this section, we first describe the approximations we are making during the analysis for the connectivity of the network, and for the spatiotemporal correlations.

After this, we show how the distribution of filtered synaptic input $h(t)$ in one subpopulation can be approximated as a Gaussian with known time-dependent parameters.

We then explain how to treat the presence of a diversity of synaptic input in one subpopulation by separating two different stochasticities: first, the intrinsic noise that makes single neurons fire stochastically in response to a given

deterministic stimulating current, and second, the presence of a stochastic input with a Gaussian distribution over a subpopulation of neurons. After averaging over the first stochasticity, we treat the correlations over the subpopulation of neurons between the values of the filtered synaptic input h_i, and the values of the adaptation variable. More specifically, we develop a simple linearization of the intrinsic-noise averaged adaptation induced by a given synaptic input.

We then show how to compute the non-linear average over the different synaptic inputs present in the network, by making use of known analytical results for the moments of the exponential of a normal variable.

The final (approximate) analytical formulas for the average rate within a subpopulation are reducible to simple non-linear differential equations.

8.2.1 Separation of the network in subpopulations

We considered cases when the neuronal network can be separated into different subpopulations, each subpopulation consisting of neurons which can be modeled by the same parameters, and which receive (resp. send), when averaged over repetitions of the same protocol, the same synaptic inputs (resp. outputs). In a realistic setting, of course, there has to be some heterogeneity between different neurons of a group, but we assume that one can find a subpopulation of neurons for which this variability can be neglected. We stress that, concerning the intrinsic dynamical properties of the neurons, a study has shown that the adaptation properties of pyramidal neurons from layer 5 of somatosensory cortex were well conserved. Indeed, a very good fit could be reached when imposing for all neurons a low-dimensional mathematical expression (a power-law) for the adaptation kernel. Further, there was only a minor variability of the three parameters of this power law between neurons [Pozzorini et al. (2013)]. Note that this result is not necessarily inconsistent with a large variability in the ion channel composition between neurons, because the absence of one ion channel may be compensated by other channels [Marder et al. (2015)]. This could in particular occur if there is some global homeostatic mechanism on the dynamical properties of the neuron.

Note that while we assume a negligibly small variance for the combined leak-and-synapse filters (and, therefore, for the synaptic weights) and for the number of connections within a subpopulation, we will briefly clarify later that the formulas

appear to be generalizable to account for more synaptic-input variability if needed.

In the following, we will use the index p to indicate one of the N_{pop} subpopulations of neurons, and the index i_p for the n_p different neurons of a given subpopulation p. We stress that some subpopulations of neurons will be recurrently connected, while some other subpopulations just provide feedforward, external stimulation to the recurrent network. The aim of the analysis is to determine the mean firing rate of recurrent populations in response to a (known) time-dependent external input coming from the external populations.

Finally, we will assume that the number of inputs from each subpopulation received by one neuron in the circuit is rather large, i.e. large enough to allow the convergence of the central limit theorem, as we will discuss later. Given that neurons in the neocortex typically receive several thousands of connections in total [Megías et al. (2001)], we are assuming that these thousands of inputs could be split in a few groups, such that the inputs within each group have similar synaptic parameters and similar firing rate modulations.

8.2.2 Assumptions about spatio-temporal correlations and their consequences

For the recurrent neurons, we assume that the spike trains are approximately emitted according to an inhomogeneous Poisson process which depends on the (dynamic) input and which is uncorrelated between units.

For the sake of clarity, we would like to take advantage of the space allowed in a Ph.D. dissertation in order to make this statement mathematically explicit with simple binary variables. Let us call $X_i(t_o)$ a variable that takes the value 1 if a neuron i from a given subpopulation fired during a time-step dt taken around time t_o, and 0 else. Note that we impose that dt is small enough such that the neuron fires at most one spike within this interval. By construction, $X_i(t_o)$ is a Bernoulli variable which expectation is dt times the rate of neuron i at time t_0. In other words,

$$prob(X_i(t_o) = 1) = [dt\ R_i(t_o)] = 1 - prob(X_i(t_o) = 0) \tag{8.10}$$

Note that $X_i(t_o)$ corresponds to the convolution of $S_i(t_o)$ with a rectangular

function $Rect_{dt}(s) = \frac{1}{dt}\left(\Theta\left(s+\frac{dt}{2}\right) - \Theta\left(s-\frac{dt}{2}\right)\right)$, where Θ is the Heaviside step function. Formally, we can write: $X_i(t_o) = (Rect_{dt} * S_i)(t_o)$.

Concerning correlations within the network for the variables X, we assume the following:

1. For any neuron i from the recurrent subpopulation, and $\forall\, t_1 \neq t_2$:

$$\begin{aligned} E_{rep\ det}\left[X_i(t_1)\ X_i(t_2)\right] &= E_{rep\ det}\left[X_i(t_1)\right]\ E_{rep\ det}\left[X_i(t_2)\right] \\ &\quad + Cov_{\ rep\ det}\left(X_i(t_1), X_i(t_2)\right) \\ &\approx E_{rep\ det}\left[X_i(t_1)\right]\ E_{rep\ det}\left[X_i(t_2)\right] \end{aligned} \tag{8.11}$$

   where $E_{rep\ det}$ is an expectation over different repetitions of the same deterministic stimulation of one neuron (i.e. the neuron is stimulated different times with the same deterministic current). In addition, $Cov_{\ rep\ det}$ is the covariance over these repetitions, which we assume to be small.

   Hence, we neglect the presence of co-occurrences between spike times that go beyond those that can be captured through a time-dependent firing rate (which is an average over different repetitions of the same deterministic current). Note, however, that the firing rate at time t_2 can still be computed as a function of the past history of firing rates at all times $t < t_2$.

   To illustrate, let us take the example of a neuron with "classical" hyperpolarizing adaptation. Qualitatively speaking, a large previous firing rate of this neuron predicts a reduced future excitability. However, this prediction is imperfect for a given trial because the specific realisation of spiking history, that directly shapes the future excitability, is only approximately matched to the past firing rates (see the EME1 approximation in [Naud et al. (2011)], and subsection 8.2.4). This approximation is expected to be rather good if the neuron is driven by a very fluctuating current which triggers almost deterministic firing at some precise times. In contrast, the very regular spike trains emitted in response to a supra-threshold constant input are much more shaped by spike time correlations.

2. For any two neurons $i \neq j$, and for all (possibly equal) times $\{t_1, t_2\}$:

$$
\begin{aligned}
E_{rep\ stoch}\left[X_i\left(t_1\right)\ X_j\left(t_2\right)\right] &= E_{rep\ stoch}\left[X_i\left(t_1\right)\right]\ E_{rep\ stoch}\left[X_j\left(t_2\right)\right] \\
&\quad + Cov_{\ rep\ stoch}\left[X_i\left(t_1\right), X_j\left(t_2\right)\right] \\
&\approx E_{rep\ stoch}\left[X_i\left(t_1\right)\right]\ E_{rep\ stoch}\left[X_j\left(t_2\right)\right] \\
\iff E_{pop\ nrn}\left[X_i\left(t_1\right)\ X_j\left(t_2\right)\right] &= E_{pop\ nrn}\left[X_i\left(t_1\right)\right]\ E_{pop\ nrn}\left[X_j\left(t_2\right)\right] \\
&\quad + Cov_{\ pop\ nrn}\left[X_i\left(t_1\right), X_j\left(t_2\right)\right] \\
&\approx E_{pop\ nrn}\left[X_i\left(t_1\right)\right]\ E_{pop\ nrn}\left[X_j\left(t_2\right)\right]
\end{aligned}
\tag{8.12}
$$

where:

- $E_{rep\ stoch}$ is an expectation over different repetitions of the stimulation of a network, such that in each repetition the stochastic external stimulation is redrawn. Hence, here, the expectation has to account for both the intrinsic stochasticity internal to each neuron, and for the variability in the synaptic input received by different neurons. Concretely, for each repetition, the spike trains coming from the external subpopulations are redrawn from a fixed random vector of time-dependent rates. $Cov_{\ rep\ stoch}$ is the covariance over these repetitions, which we assume to have a negligible effect –compared to the effect of the time-dependent firing rates– for determining the co-occurrence of spike patterns from two neurons.

- $E_{pop\ nrn}$ is the average over different neurons of the subpopulation(s) to which the neurons i and j belong. Note that they may actually belong to the same subpopulation. $Cov_{\ pop\ nrn}$ is the covariance over pairs of neurons taken from the respective subpopulation(s) to which neurons i and j belong. We assume that this covariance has a negligible effect –compared to the effect of the time-dependent firing rates– for determining the co-occurrence of spike patterns from two neurons.

We actually assume in the first part of Equation 8.12 that different neurons of a network emit spikes, and therefore receive currents, whose

joint probability is completely determined by the time-dependent stochastic-repetition-averaged firing rates, hence neglecting any correlation present in recurrent (internal) currents that would be specific to a given stimulation. This therefore implies that $E_{rep\ stoch}$ is equivalent to the average over different neurons of the concerned subpopulations (i.e., $E_{pop\ nrn}$). This equivalence holds exactly in our case where all neurons receive the same number of inputs. In a more general case, the formulas and their implications are unchanged; one would just need to account for the additional variability in synaptic input in $E_{rep\ stoch}$.

In conclusion, we neglect the effects of the correlations which arise (directly and through indirect recurrent loops) because of shared inputs between two neurons. We also neglect correlations that would arise through correlated activity between some external synapses. Finally, we neglect co-occurrences of spikes from the neurons of the recurrent population which arise through the dynamics, and which cannot be explained by two samples taken from the firing rates in the populations at the relevant times. Note that, under some biologically plausible conditions (i.e. in case of detailed balance between excitatory and inhibitory current), a recent study showed that even when these correlations were present and rather strong, their effects could effectively cancel in the total synaptic current that drives the dynamics of the neurons [Renart et al. (2007)].

We stress that the approximations in Equation 8.11 and in Equation 8.12 do not concern averages of the firing rate that would be taken over time. Hence, there can be temporal covariations of the neurons relative to their time-averaged rate [Brunel (2000); Renart et al. (2007)]. Also, we note that the expected firing rate may not only depend on the current synaptic input, but also on the previous expected firing rate history. Hence, there can be temporal correlations in the firing probability of one population, relative to its time-averaged firing rate, beyond those imposed by the synaptic input.

We now outline a few useful consequences of our assumptions about the spatiotemporal correlations. The expert reader may choose to skip those, which relate to the computation of the time-dependent expectation and variance of the sum of filtered spike trains.

1. Let us define, for any neuron i in subpopulation p_i,

$Y_i(t) = \sum_{s=0}^{\infty} \alpha_{p_i}(s) X_i(t-s)$, where $\forall s,\ \alpha_{p_i}(s) \in \mathbb{R}$ are fixed (i.e., stationary and non-random) numbers that are the same for all neurons within a subpopulation. As a first direct consequence of Equation 8.12 and of the linearity of the expectation, the variables Y are also uncorrelated between different neurons. Note that Y_i can be written as a convolution between some kernel and the spike train S_i (written as a sum of dirac deltas) that is associated with X_i. Indeed, using the previously defined rectangular filter $Rect_{dt}$, we can write:

$$\begin{aligned} Y_i(t) &:= \sum_{s=0}^{\infty} \alpha_{p_i}(s) \left[\int Rect_{dt}\left((t-s)-s'\right) S(s')\ ds' \right] \\ &= \int \left[\sum_{s=0}^{\infty} \alpha_{p_i}(s)\, Rect_{dt}\left((t-s')-s\right) \right] S(s')\ ds' \\ &= \int F(t-s')\, S(s')\ ds' := F * S(t) \end{aligned} \tag{8.13}$$

where F is a filter that is defined in continuous time.

Hence, for all (possibly equal) times t_1 and t_2, and for all neurons $i \neq j$ (but which may belong to the same subpopulation, i.e. p_i may be the same as p_j):

$$\begin{aligned} E_{pop\ nrn}\left[Y_i(t_1)\ Y_j(t_2)\right] &:= E_{pop\ nrn}\left[\sum_{s=0}^{\infty} \alpha_{p_i}(s) X_i(t_1-s) \sum_{s'=0}^{\infty} \alpha_{p_j}(s') X_j(t_2-s') \right] \\ &\iff \\ E_{pop\ nrn}\left[Y_i(t_1)\ Y_j(t_2)\right] &= \sum_{s=0}^{\infty} \sum_{s'=0}^{\infty} \alpha_{p_i}(s)\, \alpha_{p_j}(s')\, E_{pop\ nrn}\left[X_i(t_1-s)\ X_j(t_2-s')\right] \\ &\iff \\ E_{pop\ nrn}\left[Y_i(t_1)\ Y_j(t_2)\right] &\approx \sum_{s=0}^{\infty} \sum_{s'=0}^{\infty} \alpha_{p_i}(s)\, \alpha_{p_j}(s')\, E_{pop\ nrn}\left[X_i(t_1-s)\right]\ E_{pop\ nrn}\left[X_j(t_2-s')\right] \\ &\iff \\ E_{pop\ nrn}\left[Y_i(t_1)\ Y_j(t_2)\right] &\approx E_{pop\ nrn}\left[\sum_{s=0}^{\infty} \alpha_{p_i}(s) X_i(t_1-s) \right] E_{pop\ nrn}\left[\sum_{s'=0}^{\infty} \alpha_{p_j}(s') X_j(t_2-s') \right] \\ &\iff \\ E_{pop\ nrn}\left[Y_i(t_1)\ Y_j(t_2)\right] &\approx E_{pop\ nrn}\left[Y_i(t_1)\right]\ E_{pop\ nrn}\left[Y_j(t_2)\right] \end{aligned} \tag{8.14}$$

2. A similar argument (relying on the linearity of the expectation) can be made for the variables $SY_{p_1} = \sum_{i=1}^{n_{p_1}} Y_{i \in p1}(t)$ (resp. $SY_{p_2} = \sum_{i=1}^{n_{p_2}} Y_{i \in p2}(t)$) associated with two different subpopulations of neurons p_1 and p_2 which send n_{p_1} (resp. n_{p_2}) connections on a given post-synaptic target. Hence, SY_{p_1} and SY_{p_2} are uncorrelated.

3. Finally, a consequence of Equation 8.11 is an explicit expression for the variance of $Y_{i_{p_o}}(t) = \sum_{s=0}^{\infty} \alpha_{p_o}(s) X_{i_{p_o}}(t-s)$ over different neurons of one subpopulation p_o. Under our assumptions, this is equivalent to looking at the variability of a function of the response of one neuron of p_o ($Y_{i_{p_o}}(t)$), over different realisation of the stochastic time-dependent input. We recall that $Y_{i_{p_o}}(t)$ is a function that makes a weighted time-average of the variables $X_{i_{p_o}}(t)$. These variables spanning the different time steps are by construction Bernoulli variables with an expectation $(R_{p_o}(t)\ dt)$, where $R_{p_o}(t)$ is the average rate at time t over the subpopulation p_o.

   We start by noting that the variance of the sum of uncorrelated variables (in a pairwise fashion) is the sum of the variances, as well as reminding that $\forall \alpha \in \mathbb{R},\ var[\alpha\ X] = \alpha^2\ var[X]$. Hence, we can write:

$$
\begin{aligned}
&var_{pop\ nrn\ po}\left[Y_{i_{p_o}}(t)\right] := var\left[\sum_{s=0}^{\infty} \alpha_{p_o}(s) X_{i_{p_o}}(t-s)\right] \Leftrightarrow \\
&var_{pop\ nrn\ po}\left[Y_{i_{p_o}}(t)\right] \approx \sum_{s=0}^{\infty} var\left[\alpha_{p_o}(s) X_{i_{p_o}}(t-s)\right] \Leftrightarrow \\
&var_{pop\ nrn\ po}\left[Y_{i_{p_o}}(t)\right] \approx \sum_{s=0}^{\infty} \left(\alpha_{p_o}(s)\right)^2 var\left[X_{i_{p_o}}(t-s)\right] \Leftrightarrow \\
&var_{pop\ nrn\ po}\left[Y_{i_{p_o}}(t)\right] \approx \sum_{s=0}^{\infty} \left(\alpha_{p_o}(s)\right)^2 \left[\left(dt\ R_{p_o}(t-s)\right) - \left(dt\ R_{p_o}(t-s)\right)^2\right] \Leftrightarrow \\
&var_{pop\ nrn\ po}\left[Y_{i_{p_o}}(t)\right] \approx \sum_{s=0}^{\infty} \left(\alpha_{p_o}(s)\right)^2 \left[\left(dt\ R_{p_o}(t-s)\right)\right]
\end{aligned}
\tag{8.15}
$$

   where the last line holds because $X_i(t-s)$ is a Bernoulli variable with a small probability of being 1 if neurons fire irregularly and asynchronously within a subpopulation. Hence, in these conditions, for all times and for a small enough time step dt, $(R_{p_o}(t)\ dt)$ is very small, and therefore, $(R_{p_o}(t)\ dt)^2$ is negligible.

8.2.3 Characteristics of the distribution of filtered synaptic input in a neuronal subpopulation

The synaptic-input-induced voltage fluctuations experienced by neuron i_{p_o} of a subpopulation p_o can be written as:

$$h_{i_{p_o}}(t) = \sum_{p=1}^{N_{pop}} \sum_{j=1}^{n_{p,\,p_o}} F^{tot}_{p,\,p_o} * S_j^{p,\,i_{p_o}} \tag{8.16}$$

where $n_{p,\,p_o}$ is the number of neurons in subpopulation p that send projections to one neuron of subpopulation p_o, $F^{tot}_{p,\,p_o}$ is the combined leak-and-synapse filter for this specific type of synapse (see the previous subsection, and Equation 8.9), and $S_j^{p,\,i_{p_o}}$ is the spike train of the j^{th} neuron of subpopulation p sending a connection to the neuron i_{p_o} of population p_o. Note also that the subpopulation p_o is included within the external sum over subpopulations.

For any subpopulation p, $I_{i_{p_o},p} = \sum_{j=1}^{n_{p,\,p_o}} F^{tot}_{p,\,p_o} * S_j^{p,\,i_{p_o}}$ is a sum of many identically distributed and almost uncorrelated variables (see Equation 8.14). Through the central limit theorem and its generalizations (i.e. assuming that $n_{p,\,p_o}$ is large enough, and assuming that the weak correlations do not break the convergence of the sum), $I_{i_{p_o},p}$ is expected to approximatively follow a Gaussian distribution across different neurons i_{p_o} in population p_o.

In addition, for any two different subpopulations p_1 and p_2, $I_{i_{p_o},p_1}$ and $I_{i_{p_o},p_2}$ are almost uncorrelated Gaussians (see the previous subsection, item 2). We will assume a regular form for the (weak) covariations between subpopulations, hence ensuring that the different $I_{i_{p_o},p}$ are jointly normally distributed. Under these assumptions, $h_{i_{p_o}}(t)$ is also expected to approximatively follow a Gaussian distribution among different neurons i_{p_o} belonging to the subpopulation p_o. Note that it is also possible that for a given population p, the values of $I_{i_{p_o},p}$ are correlated between neurons, while still having $h_{i_{p_o}}(t)$ uncorrelated between neurons through cancellations between positive and negative correlations Renart et al. (2010). In this case, our formulas are still valid.

Note that, for a steady-state stimulation with a short and small temporal autocorrelation of the membrane-filtered synaptic input, the same argument would predict a Gaussian distribution for the values of the subthreshold potential taken at different times by the membrane of a single neuron. This has

indeed been observed during in vivo patch clamp recordings, and in detailed models of pyramidal neurons [Destexhe et al. (2003)].

Finally, we can compute the time-dependent moments of $h_{i_{p_o}}(t)$ as a function of the firing rates of neurons averaged over subpopulations.

First, we can compute $E_{pop\ nrn \in p_o}\left[h_{i_{p_o}}(t)\right]$ using the linearity of the expectation:

$$\begin{aligned} E_{pop\ nrn\ p_o}\left[h_{i_{p_o}}(t)\right] &:= E_{pop\ nrn\ p_o}\left[\sum_{p=1}^{N_{pop}} \sum_{j=1}^{n_{p,\,p_o}} F^{tot}_{p,\,p_o} * S^{p,\,i_{p_o}}_{j}\right] \\ &= \sum_{p=1}^{N_{pop}} \sum_{j=1}^{n_{p,\,p_o}} F^{tot}_{p,\,p_o} * E_{pop\ nrn \in p_o}\left[S^{p,\,i_{p_o}}_{j}\right] \\ &= \sum_{p=1}^{N_{pop}} \sum_{j=1}^{n_{p,\,p_o}} F^{tot}_{p,\,p_o} * R_p \\ &= \sum_{p=1}^{N_{pop}} n_{p,\,p_o}\ \ F^{tot}_{p,\,p_o} * R_p \end{aligned} \tag{8.17}$$

where $R_p(t) := E_{pop\ nrn\ p}\left[S_{i_p}\right]$, i.e. $R_p(t)$ is the expected firing rate within the different neurons i_p of subpopulation p at time t.

In addition, we can approximate $var_{pop\ nrn}\left[h_{i_{p_o}}(t)\right]$. First, we use the approximation of uncorrelated firing, and the fact that the variance of the sum of uncorrelated variables is the sum of their variances, Second, we use the assumption about the asynchrony and irregularity of the spiking process. The

computation follows the same steps as Equation 8.15 in the previous subsection.

$$
\begin{aligned}
var_{pop\ nrn\ p_o}\left[h_{i_{p_o}}(t)\right] &:= var_{pop\ nrn\ p_o}\left[\sum_{p=1}^{N_{pop}}\sum_{j=1}^{n_{p,\,p_o}} F^{tot}_{p,\,p_o} * S_j^{p,\,i_{p_o}}\right] \\
&\approx \sum_{p=1}^{N_{pop}}\sum_{j=1}^{n_{p,\,p_o}} var_{pop\ nrn\ p_o}\left[F^{tot}_{p,\,p_o} * S_j^{p,\,i_{p_o}}\right] \\
&\approx \sum_{p=1}^{N_{pop}}\sum_{j=1}^{n_{p,\,p_o}} FF^{tot}_{p,\,p_o} * var_{pop\ nrn\ p_o}\left[S_j^{p,\,i_{p_o}}\right], \qquad (8.18) \\
&\quad where\ \forall s,\ FF^{tot}_{p,\,p_o}(s) := \left(F^{tot}_{p,\,p_o}(s)\right)^2 \\
&\iff \\
var_{pop\ nrn\ p_o}\left[h_{i_{p_o}}(t)\right] &\approx \sum_{p=1}^{N_{pop}} n_{p,\,p_o}\ FF^{tot}_{p,\,p_o} * R_p
\end{aligned}
$$

Hence, this variance computation would be exact for inhomogeneous and uncorrelated Poisson firing with expected rates $R_p(t)$ within a subpopulation (which are our basic assumptions, see subsection 8.2.2).

We note that the central limit theorem is often quite robust to violations of its assumptions. Hence, the convergence to a normal variable may be ensured even when summing over variables that are not identically distributed (as shown through the extensions of Lyapunov and Lindeberg). In this limit, our framework may be extended to networks where the synaptic weights are drawn from a distribution that is specific to each subpopulation (while still assuming the same shape for the synaptic-and-membrane filter within a subpopulation). In addition, it may also be possible to account for stochastic synaptic transmission [Pala and Petersen (2015)]. These extensions would simply require to adjust the computation of the variance to account for the stochasticity of the synaptic weights. Hence, one would need to evaluate $var_{pop\ nrn\ p_o}\left[w_{j,i_{p_o}}\ S_j^{p,\,i_{p_o}}(t)\right]$, where $w_{j,i_{p_o}}$ is a random variable. This could be done by using the law of total variance, for instance.

Also, we stress that we could extend our approach to account for known spatiotemporal correlations within the external spike trains received by a recurrent neuron, which would only require to account for the (given) spatiotemporal covariances in Equation 8.18. Indeed, our framework only requires that different recurrent neurons can be well-enough approximated by

uncorrelated inhomogeneous poisson processes. Hence, the assumptions about correlations described in subsection 8.2.2 are more crucial for the recurrent populations of neurons, because evaluating further the correlations between recurrent spike trains would require to solve (non-linear) self-consistent equations at each time-point. This would not be an easy numerical task, and this would annihilate the efforts to reach simple equations suitable for intuitive mathematical analysis.

In conclusion, one can compute the mean and variance of the effective input $h_{i_{p_o}}(t)$ as a function of the subpopulation rates, under the assumptions made in subsection 8.2.2. We now turn to the computation of the (self-consistent) relation between this input to one neuron of the p_o subpopulation, to the expected subpopulation rate $R_{p_o}(t)$.

8.2.4 Expression of the subpopulation rate through a separation of the stochasticities due to intrinsic noise and due to synaptic input

We aim at computing (self-consistently) the expected subpopulation rate among different neurons i_{p_o} of a given (recurrently connected) subpopulation p_o: $R_{p_o}(t) := E_{pop\ nrn\ p_o}\left[S_{i_{p_o}}(t)\right]$.

We first notice that the expectation is affected by two types of variability:

- the variability in the filtered synaptic input received by different neurons. As we showed above using our assumptions, the filtered input is approximately distributed according to a Gaussian within a subpopulation of neurons, with time-dependent means and variances that depend on the subpopulations' firing rates. Mathematically, this variability can be summarized by the distribution of the (infinite) random vector $\vec{h_{p_o}} = \{h_{p_o}(t')\}_{\forall t' \leq t}$. This random vector concatenates the different random variables assigned to different times (each random variable is assigned to a given time). Each of these random variables describes a distribution across different neurons of the subpopulation of interest p_o. In the following, we will use a (slightly shorter) notation to note a particular realization of $\vec{h_{p_o}}$ for a specific neuron i_{p_o}. Rigorously, for a neuron i_{p_o}, a particular fixed realization of the input ("frozen noise") should be written: $\forall t' \leq t,\ h_{p_o}(t') = h_{i_{p_o}}(t')$, where for each t' the left-hand side is a random

variable over the subpopulation, and the right-hand side is a particular fixed realization experienced by the neuron i_{p_o}. We will use the notation $\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}$ for this particular realization of the input vector $\vec{h_{p_o}}$ received by neuron i_{p_o} during one run of the network.

- the variability of the response of one neuron subsisting even when it receives different repetitions of an identical, deterministic current $\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}$. This variability is due to the intrinsic stochasticity of the neuron. As we wrote previously, we can note the average over this variability $E_{rep\ det}$ (for the average over **det**erministic **rep**etitions). For clarity, we now use a more explicit notation for this average over the different spike trains S_i emitted by a neuron i in response to a fixed, deterministic input history [Naud and Gerstner (2012a)]. Hence, for a given neuron i_{p_o}: $E_{rep\ det}[\cdot] := E_{S_{i_{p_o}} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}[\cdot]$.

We will use the law of total expectation in order to account for these two types of variability.

Hence, by definition, for any recurrent population p_o containing the neurons with indexes i_{p_o}, the time-dependent neuron-averaged rate R_{p_o} is:

$$R_{p_o}(t_+) := E_{pop\ nrn\ p_o}\left[S_{i_{p_o}}(t_+)\right] := \lim_{dt \to 0} \frac{Prob\left(i_{p_o}\ fires\ between\ t\ and\ (t+dt)\right)}{dt} \tag{8.19}$$

We now use the definition of the firing probability given by our single neuron model (see Equation 8.1), to write:

$$R_{p_o}(t_+) = E_{pop\ nrn\ p_o}\left[\lambda_{0,\,p_o}\ \exp\left(h_{i_{p_o}}(t) + \eta_{p_o} * S_{i_{p_o}}(t)\right)\right] \tag{8.20}$$

Note that we used a "+" subscript to stress the fact that, in the last expression, the left side of the equation is caused by the right side, and hence occurs with an infinitesimal delay compared to the right side. In continuous time, this "+" subscript can actually be dropped because this delay goes to 0 (and because of the continuity and the finiteness of our expressions at all times).

To make progress, we use here the law of total expectation. We average first over the intrinsic variability of a given single unit i_{p_o} while its synaptic input history $\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}$ is fixed (and momentarily deterministic), and second over the different synaptic input histories experienced by different neurons in the

population p_o. Hence, we can write:

$$R_{p_o}(t_+) = E_{\{h_{p_o}(t')\}_{\forall t' \leq t}}\left[E_{S_{i_{p_o}} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\left[\lambda_{0,\,p_o}\ \exp\left(h_{i_{p_o}}(t) + \eta_{p_o} * S_{i_{p_o}}(t)\right)\right]\right]$$
$$R_{p_o}(t_+) = \lambda_{0,\,p_o}\ E_{\{h_{p_o}(t')\}_{\forall t' \leq t}}\left[\exp\left(h_{i_{p_o}}(t)\right) E_{S_{i_{p_o}} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\left[\exp\left(\eta_{p_o} * S_{i_{p_o}}(t)\right)\right]\right] \tag{8.21}$$

where $\{h_{p_o}(t')\}_{\forall t' \leq t}$ is the distribution of synaptic input histories within the population .

In a recent paper [Naud and Gerstner (2012a)], an analytical expression was given for the inner expectation term, for any single unit i_o: $E_{S_{i_{p_o}} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\left[\exp\left(\eta_{p_o} * S_{i_{p_o}}(t)\right)\right]$. Indeed, this term was recognized as a moment generating functional for the random point process $S_{i_{p_o}}$ representing spike trains emitted in response to a fixed input $\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}$. Therefore, the expectation can be separated in a sum which involves different correlations functions $g_n(t_1, ..., t_n)$ of order n, $\forall n \geq 1$. For increasing values of n, the g_n involve higher and higher moments of the point process (see [Van Kampen (1992)], p. 41 and more generally p. 30-44 for the use of these functions). Note that these functions are named 'correlations' because they measure how much each moment $n > 1$ deviates from independent interactions at the $(n-1)$ and lower levels.

We give here an expression of g_n for $n \leq 2$,

- $g_1(t_1) := E_{S_{i_{p_o}} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t_1}}\left[S_{i_{p_o}}(t_1)\right]$. Hence, $g_1(t_1)$ is an expectation which averages the different probabilities of spiking at t_1 that arise from different previous spiking histories occuring in response to a fixed input history $\{h_{i_{p_o}}(t')\}_{\forall t' \leq t_1}$. We note that a more formal mathematical writing for this expectation can be found in the Methods section of [Naud and Gerstner (2012a)].

- $\forall \quad t_1 \quad \neq \quad t_2$,
  $g_2(t_1, t_2) := E_{S_{i_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\left[\left(S_{i_{p_o}}(t_1) - g_1(t_1)\right)\ \left(S_{i_{p_o}}(t_2) - g_1(t_2)\right)\right]$

  Note that we must set, $\forall t_o,\ g_2(t_o, t_o) := 0$ (see [Van Kampen (1992)], p. 31, linking to p. 30-44).

Hence, using the correlation functions g_n, one can write the following expansion:

$$E_{S_{i_{p_o}} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}} \left[\exp\left(\eta_{p_o} * S_{i_{p_o}}(t)\right)\right] =$$
$$\exp\left(\sum_{n=1}^{\infty}\left(\frac{1}{n!}\int_{-\infty}^{t}...\int_{-\infty}^{t}\left(e^{\eta_{p_o}(t-s_1)}-1\right)\ ...\ \left(e^{\eta_{p_o}(t-s_n)}-1\right)\ g_n\left(t_1,\ ...,\ t_n\right)ds_1\ ...ds_n\right)\right)$$
$$\Rightarrow$$

$$E_{S_{i_{p_o}} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}} \left[\exp\left(\eta_{p_o} * S_{i_{p_o}}(t)\right)\right] \approx \exp\left(\int_{-\infty}^{t}\left(e^{\eta_{p_o}(t-s)}-1\right)g_1(s)\ ds\right) \tag{8.22}$$

Note that the last formula, which only accounts for the first order ($n = 1$) of the expansion, would be exact if the point process was truly inhomogeneous Poisson [Van Kampen (1992) p. 33, also in Stratanovitch (1963)], i.e. if the spiking process could be completely described by a time-dependent, but history-of-firing independent, firing probability. Indeed, in this case, for any $n > 1$, all the correlation functions g_n vanish. Therefore, when using the approximation of Equation 8.22, the error will grow with the amplitude of the correlations within the spike trains, and thus, for our purpose, with the strength of the adaptation. More specifically, adaptation most often creates negative correlations between spike times [Farkhooi et al. (2011); Schwalger and Lindner (2013)], which leads to expecting negative values for g_2. Given that $\forall s,\ \eta_{p_o}(s) \leq 0$ (hence implementing a "classical" adaptation which drives the excitability down upon each spike [Pozzorini et al. (2013)]), the second order term in the sum is expected to be negative, i.e. we expect $\left(\int_{-\infty}^{t}\int_{-\infty}^{t}\left(e^{\eta_{p_o}(t-s_1)}-1\right)\left(e^{\eta_{p_o}(t-s_2)}-1\right)g_2(s_1,s_2)\,ds_1 ds_2\right) < 0$. By neglecting this second order term, we therefore conjecture to underestimate the self-inhibition coming from the adaptation variable, and hence to overestimate the predicted firing rate. This overestimation error would also be expected to grow with the firing rate, as the spike time correlations become larger for smaller interspike intervals with a realistic power-law-like adaptation kernel. This is indeed what was shown to happen in the original publication making use of this moment-based expansion ([Naud and Gerstner (2012a)], see Figure 8.1). For clarity, we reproduce here one figure adapted from this publication in order to illustrate how well the approximation in Equation 8.22 works for predicting the firing rate in response to a fixed, deterministic input.

As can be seen in Figure 8.1, the approximation based on the first moment only

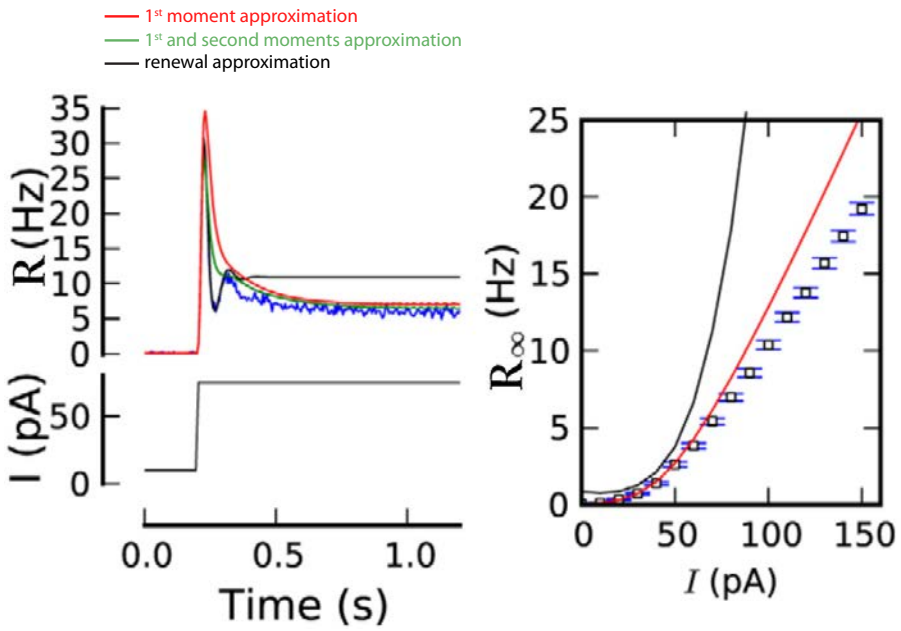


Figure 8.1: *Performance of the approximation of adaptation through the* 1^{st} *moment with a deterministic current.* Adapted from [Naud and Gerstner (2012a)]. In the figure, simulations (25 000 repetitions of the same deterministic current, blue line) are compared to theory. The single neuron model is identical to ours, and possesses a power-law-like adaptation kernel. The red line corresponding to the 1^{st} moment approximation (g_1 only, see the bottom of Equation 8.22); while the green line takes into account two moments (g_1 and g_2 in the sum in the first line of Equation 8.22). Finally, the black line approximates $\eta_{p_o} * S_{i_{p_o}}(t) \approx \eta_{p_o}(t - t_{last})$, where t_{last} is the time of the last spike fired by the neuron (i.e., it makes a renewal approximation). **Left**: time-course of the firing rate in the simulations, and comparison with the theories, in response to a deterministic step current (bottom). **Right**: steady-state firing rate in the simulation and prediction from the theories, in response to different values of a constant depolarizing current.

(g_1, red line) indeed leads to an overestimation of the firing rates, and more-so for higher rates, with an error that is slightly decreased when also accounting for the second moment (g_1 and g_2 , green line). However, this first moment approximation still accounts rather well for the time-dependent effect of adaptation on the firing rate, allowing to capture the initial peak of firing rate and to approximate the following decay of activity, while missing only minor oscillations of the rate. Also, the g_1 approximation does account for the summation occurring over different spike times (as it gives a better fit for the steady-state firing rate than a theory which only accounts for the effect of the last spike, black line in Figure 8.1).

Hence, the approximation based on g_1 is very simple while still capturing major features of the adaptation effects.

Finally, we can rewrite this approximate 1^{st} moment formula in Equation 8.22:

$$E_{S_{i_{p_o}} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}} \left[\exp\left(\eta_{p_o} * S_{i_{p_o}}(t)\right)\right] \approx \exp\left(\left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)\right) \tag{8.23}$$

where $\forall s > 0,\ \breve{\eta}_{p_o}(s) := \left(e^{\eta_{p_o}(s)} - 1\right)$ while $\forall s \leq 0,\ \breve{\eta}_{p_o}(s) := 0$, and $r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}(t) := g_1(t)$ is the average rate of a single neuron i_{p_o} over different repetitions of a stimulation with a fixed deterministic input history $\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}$.

Together with Equation 8.21, this leads us to a (non-explicit) equation for the expected population rate over different neurons receiving different inputs $R_{p_o}(t)$:

$$\begin{aligned} R_{p_o}(t_+) &\approx \lambda_{0,\,p_o}\, E_{\{h_{p_o}(t')\}_{\forall t' \leq t}} \left[\exp\left(h_{i_{p_o}}(t)\right) \exp\left(\left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)\right)\right] \\ &\approx \lambda_{0,\,p_o}\, E_{\{h_{p_o}(t')\}_{\forall t' \leq t}} \left[\exp\left(h_{i_{p_o}}(t) + \left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)\right)\right] \end{aligned} \tag{8.24}$$

However, evaluating the remaining expectation is not trivial, first because we do not know the distribution of $\left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)$ over the different neurons of the subpopulation p_o, and second because $h_{i_{p_o}}$ and $\left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)$ are strongly (negatively) correlated among this subpopulation of neurons. Indeed, the filtered input $h_{i_{p_o}}$ received by a neuron i_{p_o} is correlated over time; hence, if it takes large values at times t, it probably also took large values in the past, leading to a stronger firing in the past and

therefore to a more negative expected adaptation variable $\left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)$.

The following section describes the method we developped to get around these difficulties.

8.2.5 Explicit expression of the subpopulation rate through a linearization of the expected adaptation variable

In order to make progress from Equation 8.24, we would like to determine the distribution of $G_{i_{p_o}}(t) := \left(h_{i_{p_o}}(t) + \left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)\right)$ over the neurons i_{p_o} of the population p_o.

One way to do this is to notice that the (deterministic-repetitions averaged) adaptation variable $\left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)$ should be a function of the history of input $\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}$. If one could linearize this variable, i.e. if one could find a kernel Γ_{p_o} and a constant C_{p_o} such that:

$$\left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t) \approx \left(\Gamma_{p_o} * h_{i_{p_o}}\right)(t) + C_{p_o} \tag{8.25}$$

then, the distribution of $G_{i_{p_o}}(t)$ within the subpopulation p_o could be approximated from the distributions of $\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}$ (which were determined to be Gaussian in subsection 8.2.3).

We now turn to deriving Γ_{p_o} and C_{p_o}, by using a linearization of $\left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)$, which in turn requires to linearize $r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}$. Hence, we need a "baseline" value $R_{p_o,\, bsln}$ for the population rate, around which we can compute deviations of $r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}$.

Self-consistent derivation of the recurrent baseline firing rates

To find a value for this baseline firing rate $R_{p_o,\, bsln}$, we will make use of an "ideal network", whose firing rates can be expected to be rather similar (but not identical) to the time-averaged firing rates which occur within our more complex network of interest.

The reason for using such an approximate "baseline" firing rate in this "ideal network" is that, under these simplified conditions, we can self-consistently express these baseline subpopulation rates $R_{p\,recc,\,bsln}$ for all recurrently connected subpopulations through coupled transcendental equations. In addition, in case of a single recurrent population, or when there is only one subpopulation for which the dynamics is substantially non-linear, the equations only involve the Lambert-W function which has well-defined solutions. Hence, in this framework, there is no need to use very complex numerical recipes to solve the baseline steady-state self-consistently. In addition, we stress that we will later be able to use the baseline rates as parameters for intermediate computations which will ultimately lead to approximate the time-averaged firing rate in our more complex network.

Note that a more complex numerical approach that does not make use of the "ideal network" (and which would therefore be expected to give more accurate results, but which has the disadvantage of diminishing the mathematical tractability of our framework) may still be used to compute a mean firing rate self-consistently. A linearization around this mean firing rate would indeed be expected to minimize the error for the predicted firing rate. This approach would just require to numerically search a self-consistent match between an (initially unknown and initiated with a "guess") mean firing rate used for the linearization, and the value of the mean firing rate that our formulas provide at the end (see Equation 8.42).

Hence, we choose as a baseline the steady-state subpopulation rate in a "frozen" network receiving the same mean external synaptic drive as our original network. However, in this "frozen" network, all fluctuations are neglected (i.e. we neglect all the fluctuations of the filtered synaptic input, both external and recurrent). Such an equivalent "frozen" network can be (theoretically) constructed by making the size N_p of each subpopulation p go to infinity, while rescaling the synaptic weights from population p by N_p. Briefly, in the original network, if we take $w^{real}_{p_1,p_2}$ the synaptic weight from a neuron of subpopulation 1 to neuron of subpopulation 2 and $N^{real}_{p_1,p_2}$ as the number of neurons from p_1 projecting to a neuron of p_2, then the mean synaptic current from p_1 to one neuron of p_2 will scale as $\left(w^{real}_{p_1,p_2}\, N^{real}_{p_1,p_2}\right)$. In the "ideal network" with a number of neurons $N^{ideal}_{p_1,p_2} \to \infty$ coming from p_1 and projecting to one neuron of p_2, we can choose $w^{ideal}_{p_1,p_2} := \frac{w^{real}_{p_1,p_2}\, N^{real}_{p_1,p_2}}{N^{ideal}_{p_1,p_2}}$. Hence, the mean synaptic current from p_1 to one neuron of p_2 still scales as $\sum_{i=1}^{N^{ideal}_{p_1,p_2}} w^{ideal}_{p_1,p_2} =$

$w^{real}_{p_1,p_2}\ N^{real}_{p_1,p_2}$. However, the variance of this synaptic current will now scale as $\sum_{i=1}^{N^{ideal}_{p_1,p_2}} \left(w^{ideal}_{p_1,p_2}\right)^2 = \frac{\left(w^{real}_{p_1,p_2}\ N^{real}_{p_1,p_2}\right)^2}{N^{ideal}_{p_1,p_2}}$, which goes to 0 as $N^{ideal}_{p_1,p_2}$ goes to infinity.

Hence, for any recurrent subpopulation p_o within this "frozen" network in steady-state, all neurons receive the same baseline constant input $h_{o,\,bsln}$, which is related to the neuron-averaged h_o of the original network. In addition, now, the firing rates of the recurrent populations $R_{p_{recc},\,bsln}$ are constant and respond to constant stimulations from external subpopulations (with rates $R_{p_{ext},\,bsln} := E_t[R_{p_{ext}}(t)]$, i.e. we take the time-average of the external subpopulation rates from the original network).

As a consequence, by separating the subpopulations between recurrent (whose rates have to be determined self-consistently) and external ones, we can write:

$$\begin{aligned} h_{p_o,\,bsln} &:= \sum_{p_{recc}=1}^{N_{pop\ recc}} n_{p_{recc},\,p_o}\ F^{tot}_{p_{recc},\,p_o} * R_{p_{recc},\,bsln} + \\ &\sum_{p_{ext}=1}^{N_{pop\ ext}} n_{p_{ext},\,p_o}\ F^{tot}_{p_{ext},\,p_o} * R_{p_{ext},\,bsln} \\ &= \sum_{p_{recc}=1}^{N_{pop\ recc}} n_{p_{recc},\,p_o}\ R_{p_{recc},\,bsln}\ \left(\int F^{tot}_{p_{recc},\,p_o}\right) + \\ &\sum_{p_{ext}=1}^{N_{pop\ ext}} n_{p_{ext},\,p_o}\ R_{p_{ext},\,bsln}\ \left(\int F^{tot}_{p_{ext},\,p_o}\right) \end{aligned} \tag{8.26}$$

Given these time-independent and neuron-independent (within a subpopulation) filtered inputs, we can re-work Equation 8.24 for any recurrent population p_o:

$$
\begin{aligned}
R_{p_o,\,bsln} &\approx \lambda_{0,\,p_o}\; E_{\{h_{p_o}(t')\}_{\forall t' \leq t}}\left[\exp\left(h_{i_{p_o}}(t) + \left(\breve{\eta}_{p_o} * r_{i_{p_o}\,|\,\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)\right)\right] \\
&\approx \lambda_{0,\,p_o}\; \exp\left(h_{p_o\,bsln} + (\breve{\eta}_{p_o} * R_{p_o,\,bsln})\right) \\
&\approx \lambda_{0,\,p_o}\; \exp\left(\sum_{p_{ext}=1}^{N_{pop\;ext}} n_{p_{ext},\,p_o}\; R_{p_{ext},\,bsln}\; \left(\int F^{tot}_{p_{ext},\,p_o}\right)\right) \\
&\exp\left(R_{p_o,\,bsln}\left(n_{p_o,\,p_o}\int F^{tot}_{p_o,\,p_o} + \int \breve{\eta}_{p_o}\right) + \sum_{p_{recc}\;\neq p_o} n_{p_{recc},\,p_o}\; R_{p_{recc},\,bsln}\; \left(\int F^{tot}_{p_{recc},\,p_o}\right)\right)
\end{aligned}
\tag{8.27}
$$

This defines coupled transcendental equations for the recurrent baseline rates $R_{p_{recc},\,bsln}$, as announced previously. More precisely if we note $\vec{R}_{p_{recc},\,bsln}$ a column vector of these rates, then we can compute the values of a real column vector $\vec{C}$ and of a real matrix $\vec{M}$ such that:

$$
\vec{R}_{p_{recc},\,bsln} = \vec{C} \cdot \exp\left(\vec{M}\vec{R}_{p_{recc},\,bsln}\right) \tag{8.28}
$$

Linearization of the exponential non-linearity around the baseline firing rates

We now turn to the linearization of $r_{i_{p_o}\,|\,\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}$ around $h_{p_o,\,bsln}$, which will simply rely on a Taylor expansion for the exponential. We will write, $\forall t,\;\; \Delta h_{i_{p_o}}(t) := h_{i_{p_o}}(t) - h_{p_o,\,bsln}$, where $h_{p_o,\,bsln}$ is defined in Equation 8.26. Similarly, we will take $\Delta r_{i_{p_o}\,|\,\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}(t) := r_{i_{p_o}\,|\,\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}(t) - R_{p_o,\,bsln}$.

Starting again from the definition of $r_{i_o\,|\,\{h_{i_o}(t')\}_{\forall t' \leq t}}(t)$ as an average of the different spiking probabilities at time t arising through different spiking histories

in response to the input $\{h_{i_o}(t')\}_{\forall t' \leq t}$, we can write:

$$\begin{aligned} r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}(t) &:= E_{S_{i_{p_o}} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\left[S_{i_{p_o}}(t)\right] \\ &:= \lim_{dt \to 0} \frac{Prob\left((i_{p_o}\ fires\ at\ t) \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}\right)}{dt} \\ &:= E_{S_{i_{p_o}} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\left[\lambda_{0,\,p_o}\ \exp\left(h_{i_{p_o}}(t) + \eta_{p_o} * S_{i_{p_o}}(t)\right)\right] \\ &= \lambda_{0,\,p_o}\ \exp\left(h_{i_{p_o}}(t)\right)\ E_{S_{i_{p_o}} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\left[\ \exp\left(\eta_{p_o} * S_{i_{p_o}}(t)\right)\right] \\ &\approx \lambda_{0,\,p_o}\ \exp\left(h_{i_{p_o}}(t)\right)\ \exp\left(\left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)\right) \end{aligned} \tag{8.29}$$

where we used the result from Equation 8.23 for the last step.

From this last equation, we will now use $\Delta h_{i_{p_o}}$ and $\Delta r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}$ to write:

$$\begin{aligned} &\lambda_{0,\,p_o}\ \exp\left(h_{i_{p_o}}(t) + \left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)\right) \\ &\quad = \lambda_{0,\,p_o} \exp\left(\Delta h_{i_{p_o}}(t) + h_{p_o,\,bsln} + \left(\breve{\eta}_{p_o} * \left(\Delta r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}} + R_{p_o,\,bsln}\right)\right)(t)\right) \\ &\quad = \lambda_{0,\,p_o}\ \exp\left(h_{p_o,\,bsln} + (\breve{\eta}_{p_o} * R_{p_o,\,bsln})\right) \exp\left(\Delta h_{i_{p_o}}(t) + \left(\breve{\eta}_{p_o} * \Delta r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)\right) \\ &\quad = R_{p_o,\,bsln}\ \exp\left(\Delta h_{i_{p_o}}(t) + \left(\breve{\eta}_{p_o} * \Delta r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)\right) \end{aligned} \tag{8.30}$$

Where we used the second line of Equation 8.27 in the last step.

We will now Taylor-expand the exponential for small $\Delta Exc_{\,i_{p_o}}(t) := \left(\Delta h_{i_{p_o}} + \left(\breve{\eta}_{p_o} * \Delta r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)\right)(t)$. Hence, we will take $\exp\left(\Delta Exc_{\,i_{p_o}}(t)\right) = 1 + \Delta Exc_{\,i_{p_o}}(t) + \epsilon\left(\Delta Exc_{\,i_{p_o}}(t)\right)$. Note that the error term (ϵ) accounts for all higher-order terms. We will keep this remaining total error explicitly in the formula for now, and we will see later how to approximately account for it.

We note that $\Delta h_{i_{p_o}}$ and $\left(\breve{\eta}_{p_o} * \Delta r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)$ are negatively correlated, which is favorable for the linearization as it will tend to bring $\Delta Exc_{\,i_{p_o}}$ close to 0.

To first order, this linearization will actually lead to an underestimation of $r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}(t)$ as, $\forall x,\ exp(x) \geq (1 + x)$. We can actually quantify the peformance of this linearization to first order as a function of the ratio

$\frac{r_{i_{po} \mid \{h_{i_{po}}(t')\}_{\forall t' \leq t}}}{R_{p_o,\, bsln}}$. If this ratio is 2, the error ϵ is $2 - 1 - \ln(2) \approx 0.31$ (i.e. we estimate a rate $r_{i_{po} \mid \{h_{i_{po}}(t')\}_{\forall t' \leq t}}$ of $\approx 1.7\ R_{p_o,\, bsln}$ instead of the actual value of $2\ R_{o,\, bsln}$). Similarly, if the ratio is 0.5, $\epsilon \approx 0.5 - 1 - \ln(0.5) \approx 0.19$ (i.e. we estimate a rate $r_{i_{po} \mid \{h_{i_{po}}(t')\}_{\forall t' \leq t}}$ of $0.3\ R_{p_o,\, bsln}$ instead of the true value of $0.5\ R_{p_o,\, bsln}$).

Note that this approach is much simpler than the only other currently available linearization procedure for this type of adapting neuron model [Deger et al. (2014)]. This simplification was permitted by the decision to start from an expression for the intrinsic-stochasticity averaged adaptation which does not differentiate the last spike from the previous ones, and treats the whole spiking history through a first-moment approximation (see Equation 8.22). Instead, Deger et al. (2014) use an additional integral over the time of the last spike, in order to more accurately account for possibly strong refractory effects. We note that actually, in Deger et al. (2014), the evaluated linear kernel was semi-analytical as it required the steady-state interspike interval distribution, which was taken from the simulation. In contrast, we will derive and evaluate a fully analytical expression, as we demonstrate below.

Hence, using this approximation, we can write:

$$r_{i_{po} \mid \{h_{i_{po}}(t')\}_{\forall t' \leq t}} \approx R_{p_o,\, bsln} \left(1 + \left(\Delta h_{i_{po}} + \left(\breve{\eta}_{p_o} * \Delta r_{i_{po} \mid \{h_{i_{po}}(t')\}_{\forall t' \leq t}}\right)\right) + \epsilon \left(\Delta Exc_{\, i_{po}}\right)\right)$$

$$\Leftrightarrow$$

$$\Delta r_{i_{po} \mid \{h_{i_{po}}(t')\}_{\forall t' \leq t}} \approx R_{p_o,\, bsln} \left(\Delta h_{i_{po}} + \left(\breve{\eta}_{p_o} * \Delta r_{i_{po} \mid \{h_{i_{po}}(t')\}_{\forall t' \leq t}}\right) + \epsilon \left(\Delta Exc_{\, i_{po}}\right)\right) \tag{8.31}$$

We now divide by $R_{p_o,\, bsln}$, and collect the terms that are linear in $\Delta r_{i_{po} \mid \{h_{i_{po}}(t')\}_{\forall t' \leq t}}$:

$$\textit{Equation 8.31} \Leftrightarrow \left(\frac{\delta}{R_{p_o,\, bsln}} - \breve{\eta}_{p_o}\right) * \Delta r_{i_{po} \mid \{h_{i_{po}}(t')\}_{\forall t' \leq t}} \approx \left(\Delta h_{i_{po}}\right) + \epsilon \left(\Delta Exc_{\, i_{po}}\right) \tag{8.32}$$

where δ denotes the Dirac δ distribution.

Taking the Fourier transform $\mathfrak{F}[\cdot]$:

Equation 8.32 $\Leftrightarrow$

$$\mathfrak{F}\left[\frac{\delta}{R_{p_o,\,bsln}} - \breve{\eta}_{p_o}\right](s)\;\; \mathfrak{F}\left[\Delta r_{i_{p_o}\,|\,\{h_{i_{p_o}}(t')\}_{\forall t'\leq t}}\right](s) \approx \mathfrak{F}\left[\Delta h_{i_{p_o}}\right](s) + \mathfrak{F}\left[\epsilon\left(\Delta Exc_{\,i_{p_o}}\right)\right](s)$$

$\Leftrightarrow$

$$\mathfrak{F}\left[\Delta r_{i_{p_o}\,|\,\{h_{i_{p_o}}(t')\}_{\forall t'\leq t}}\right](s) \approx \frac{1}{\mathfrak{F}\left[\frac{\delta}{R_{p_o,\,bsln}} - \breve{\eta}_{p_o}\right](s)}\,\mathfrak{F}\left[\Delta h_{i_{p_o}}\right](s) + \frac{1}{\mathfrak{F}\left[\frac{\delta}{R_{p_o,\,bsln}} - \breve{\eta}_{p_o}\right](s)}\,\mathfrak{F}\left[\epsilon\left(\Delta Exc_{\,i_{p_o}}\right)\right](s) \tag{8.33}$$

Finally, taking the inverse Fourier transform $\mathfrak{F}^{-1}$ and defining $\Lambda_{p_o} := \mathfrak{F}^{-1}\left[\frac{1}{\mathfrak{F}\left[\frac{\delta}{R_{p_o,\,bsln}} - \breve{\eta}_{p_o}\right]}\right]$, we find:

Equation 8.33 $\Leftrightarrow$

$$\Delta r_{i_{p_o}\,|\,\{h_{i_{p_o}}(t')\}_{\forall t'\leq t}}(t) = \left(\Lambda_{p_o} * \Delta h_{i_{p_o}}\right)(t) + \left(\Lambda_{p_o} * \epsilon\left(\Delta Exc_{\,i_{p_o}}\right)\right)(t)$$

$\Leftrightarrow$

$$r_{i_{p_o}\,|\,\{h_{i_{p_o}}(t')\}_{\forall t'\leq t}}(t) \approx \left(\Lambda_{p_o} * h_{i_{p_o}}\right)(t) + \left(\Lambda_{p_o} * \epsilon\left(\Delta Exc_{\,i_{p_o}}\right)\right)(t) + \left(R_{p_o,\,bsln} - h_{p_o,\,bsln}\int \Lambda_{p_o}\right) \tag{8.34}$$

Evaluating the error term $\left(\Lambda_{p_o} * \epsilon\left(\Delta Exc_{\,i_{p_o}}\right)\right)(t)$ is in general difficult, as it requires to compute self-consistently higher moments of the rate distribution. This error term is time-dependent, and in addition it has a bias (i.e., its average is non-zero). Indeed, we explained previously why the linearization to first order leads to a systematic underestimation of the firing rate, and thus an error term ϵ that is always positive. In consequence, $\left(\Lambda_{p_o} * \epsilon\left(\Delta Exc_{\,i_{p_o}}\right)\right)(t)$ is always positive, with larger values when the time-dependent rate $r_{i_{p_o}\,|\,\{h_{i_{p_o}}(t')\}_{\forall t'\leq t}}$ of the neuron i_{p_o} gets further away from $R_{p_o,\,bsln}$. In addition, we remind that we will at the end only be interested in a linear approximation for the average adaptation $\left(\breve{\eta}_{p_o} * r_{i_{p_o}\,|\,\{h_{i_{p_o}}(t')\}_{\forall t'\leq t}}\right)(t)$. Using Equation 8.34, this gives

$$\left(\breve{\eta}_{p_o} * \Lambda_{p_o} * h_{i_{p_o}}\right) + \left(\int \breve{\eta}_{p_o}\right)\left(R_{p_o,\,bsln} - h_{p_o,\,bsln}\int \Lambda_{p_o}\right) + \left(\breve{\eta}_{p_o} * \Lambda_{p_o} * \epsilon\left(\Delta Exc_{\,i_{p_o}}\right)\right) \tag{8.35}$$

Hence, concerning this linear approximation for the adaptation variable, for each single neuron with a time-dependent firing induced by a fluctuating fixed input history $\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}$, the (biased) error term gets low-pass filtered and "averaged" by the kernel $\breve{\eta}_{p_o}$.

This suggests that the inaccuracies of this linear prediction could be well-reduced through an approximation of the time-dependent error by a constant, hence correcting for the average bias. We will compute this constant correction term from an estimation of the average error ϵ occurring during the deviations of the firing rate from $R_{p_o,\, bsln}$ while the network is in a fluctuating steady-state. The fluctuations during this steady-state indeed arise because different neurons receive different synaptic inputs. More precisely, in this steady-state, the external populations feed the recurrent ones with spike trains that have a rate $R_{p_{ext},\, ss} := E_t[R_{p_{ext}}(t)]$, that is thus constant over time. This correction term will be computed self-consistently later (in Equation 8.46).

To illustrate, we would like to describe the effect of this correction term during the above-mentioned steady-state regime. In this regime, at a given time, different neurons of p_o fire at different rates that deviate with different magnitudes from $R_{p_o,\, bsln}$. In this context, the correction will make the error approximately homogeneous among different neurons with different deviations from $R_{p_o,\, bsln}$. More precisely, those neurons close to $R_{p_o,\, bsln}$ will have a slightly overestimated firing rate through the corrected linearization, while those neurons further away from $R_{p_o,\, bsln}$ will have a slightly underestimated firing rate through the corrected linearization. This differs from the consistent underestimation of the firing rate when using the uncorrected formula. Finally, this approximate correction will be important when we will use Equation 8.34 to estimate the adaptation term $\left(\breve{\eta}_{p_o} * r_{i_{p_o} \,|\, \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)$ entering in the (non-linear) expression for the single-neuron firing probability (see Equation 8.24 and Equation 8.39 below).

Hence, we define the following time-independent correction term for the linearized intrinsic-stochasticity-averaged adaptation, using the above-mentioned fluctuating steady-state regime (corresponding to the index ss):

$$\begin{aligned} A_{p_o,\, fluct} :&= E_{\{h_{p_o}(t')\}_{\forall t' \leq t} \, during \, ss}\left[\breve{\eta}_{p_o} * \left(\Lambda_{p_o} * \epsilon\left(\Delta Exc_{\, i_{p_o},\, ss}\right)\right)\right] \\ &= \left(\int \breve{\eta}_{p_o}\right) E_{\{h_{p_o}(t')\}_{\forall t' \leq t} \, during \, ss}\left[\Lambda_{p_o} * \epsilon\left(\Delta Exc_{\, i_{p_o},\, ss}\right)\right] \end{aligned} \tag{8.36}$$

This allows us to express a linearized formula for the time-dependent rate that includes the (time-independent) correction term:

$$r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}(t) \approx \left(\Lambda_{p_o} * h_{i_{p_o}}\right)(t) + \left(R_{p_o,\, bsln} + \frac{A_{p_o,\, fluct}}{\int \breve{\eta}_{p_o}} - h_{p_o,\, bsln} \int \Lambda_{p_o}\right) \tag{8.37}$$

Finally, we get the desired result: a linearized equation for the intrinsic-stochasticity averaged adaptation in response to a fixed input history $\{h_{i_{p_o}}(t')\}_{\forall t' \leq t}$:

$$\left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t) \approx \left(\Gamma_{p_o} * h_{i_{p_o}}\right)(t) + C_{p_o} \tag{8.38}$$

Where $\Gamma_{p_o} := \breve{\eta}_{p_o} * \Lambda_{p_o}$, and $C_{p_o} := A_{p_o,\, fluct} + \left(\int \breve{\eta}_{p_o}\right)\ \left(R_{p_o,\, bsln} - h_{p_o,\, bsln}\ \int \Lambda_{p_o}\right)$.

We will now be able to compute the firing rates of the recurrent populations in the steady-state self-consistently while accounting for the heterogeneity of input over neurons (whereas the baseline firing rates computed above neglected all fluctuations and the heterogeneity of input over neurons).

Self-consistent computation of the steady-state recurrent firing rates

We can finally express the expected rate of a neuron within a subpopulation p_o: R_{p_o}, which is an average over the different synaptic inputs received by the different neurons of p_o. Hence, from Equation 8.24, Equation 8.16 and Equation 8.38, we can write:

$$\begin{aligned}
R_{p_o}(t) &\approx \lambda_{0,\, p_o}\ E_{\{h_{p_o}(t')\}_{\forall t' \leq t}}\left[\exp\left(h_{i_{p_o}}(t) + \left(\breve{\eta}_{p_o} * r_{i_{p_o} \mid \{h_{i_{p_o}}(t')\}_{\forall t' \leq t}}\right)(t)\right)\right] \\
&\approx \lambda_{0,\, p_o}\ E_{\{h_{p_o}(t')\}_{\forall t' \leq t}}\left[\exp\left(h_{i_{p_o}}(t) + \left(\Gamma_{p_o} * h_{i_{p_o}}\right)(t) + C_{p_o}\right)\right] \\
&\approx \lambda_{0,\, p_o}\ \exp\left(C_{p_o}\right)\ E_{\{h_{p_o}(t')\}_{\forall t' \leq t}}\left[\exp\left(\sum_{p=1}^{N_{pop}} \sum_{j=1}^{n_{p,\, p_o}} \Phi_{p,\, p_o} * S_j^{p,\, i_{p_o}}\right)\right] \\
&\text{where } \Phi_{p,\, p_o} := \left(F_{p,\, p_o}^{tot} + \left(\Gamma_{p_o} * F_{p,\, p_o}^{tot}\right)\right)
\end{aligned} \tag{8.39}$$

Through the same argument as in subsection 8.2.3,

$$Z_{i_{p_o}} = \sum_{p=1}^{N_{pop}} \sum_{j=1}^{n_{p,\, p_o}} \Phi_{p,\, p_o} * S_j^{p,\, i_{p_o}} = h_{i_{p_o}} + \left(\Gamma_{p_o} * h_{i_{p_o}}\right) \tag{8.40}$$

should converge to a Gaussian variable for a large enough number of inputs. In addition, compared to the variable h_{i_o} in subsection 8.2.3, we just replaced the kernels F^{tot} by the kernels Φ. Hence, we can immediately deduce that:

$$\begin{aligned} E_{pop\ nrn\ p_o}\left[Z_{i_{p_o}}(t)\right] &= \sum_{p=1}^{N_{pop}} n_{p,\,p_o} \quad \left(\Phi_{p,\,p_o} * R_p\right)(t) \\ var_{pop\ nrn\ p_o}\left[Z_{i_{p_o}}(t)\right] &\approx \sum_{p=1}^{N_{pop}} n_{p,\,p_o} \Phi\Phi_{p,\,p_o} * R_p\,(t) \\ &\quad where\ \forall s,\ \Phi\Phi_{p,\,p_o}(s) := \left(\Phi_{p,\,p_o}(s)\right)^2 \end{aligned} \tag{8.41}$$

These results imply that $\exp\left(Z_{i_{p_o}}(t)\right)$ is a log-normal variable, whose expectation only depends on the mean and variance of X_{i_o}. Hence, at any time:

$$R_{p_o}(t) \approx \lambda_{0,\,p_o}\ \exp\left(C_{p_o}\right)\ \exp\left(E_{pop\ nrn\ p_o}\left[Z_{i_{p_o}}(t)\right] + \frac{var_{pop\ nrn\ p_o}\left[Z_{i_{p_o}}(t)\right]}{2}\right) \tag{8.42}$$

which is the desired result, namely the expected average firing rate in the recurrent populations of neurons. Note that here, the effects of the synaptic input variance are considered.

Finally, we can express the steady-state firing rate of a recurrent population p_o, and determine $A_{p_o,\,fluct}$ (which enters in the definition of C_{p_o}, see Equation 8.38) self-consistently. For the recurrent populations, this steady-state is different from the baseline steady-state considered above (in section 8.2.5) as the effect of the synaptic input variance will not be neglected.

We split again the subpopulations into external and recurrent ones. Their population firing rates are constant over time, and will be written $R_{p_{ext},\,ss}$ and $R_{p_{recc},\,ss}$, respectively. Hence, decorating all steady-state quantity with ss:

$$E_{pop\ nrn\ p_o}\left[Z_{i_{p_o},ss}\right] \approx \sum_{p=1}^{N_{pop}} n_{p,\,p_o}\ \ R_{p,\,ss}\ \left(\int \Phi_{p,\,p_o}\right)$$

$$var_{pop\ nrn\ p_o}\left[Z_{i_{p_o},ss}\right] \approx \sum_{p=1}^{N_{pop}} n_{p,\,p_o}\ \ R_{p,\,ss}\ \left(\int \Phi\Phi_{p,\,p_o}\right)$$

$$R_{p_o,\,ss} \approx \lambda_{0,\,p_o}\ \exp\left(C_{p_o} + \left(\sum_{p_{ext}} n_{p_{ext}} R_{p_{ext},\,ss}\left(\int \Phi_{p_{ext},\,p_o} + 0.5\int \Phi\Phi_{p_{ext},\,p_o}\right)\right)\right)$$
$$\exp\left(\sum_{p_{recc}} n_{p_{recc}} R_{p_{recc},\,ss}\left(\int \Phi_{p_{recc},\,p_o} + 0.5\int \Phi\Phi_{p_{recc},\,p_o}\right)\right) \tag{8.43}$$

This again defines coupled transcendental equations between recurrent populations, which reduce to the Lambert-W function for a single recurrent population.

Note that there would be a possibility to try to Taylor-expand the exponential again, which would yield linearized rate equations with a dependence on both a mean, and a variance, synaptic drive.

Finally, one can define the $A_{p,\,fluct}$ self-consistently by recomputing the expected firing rate over the population from the linearized formula giving the average firing rate over the intrinsic stochasticity. We consider a steady-state during which the external subpopulations fire at a rate that is a time-average of the rate they have during the (possibly time-dependent) stimulation. In other words, $R_{p_{ext},\,ss} := E_t[R_{p_{ext}}(t)] = R_{p_{ext},\,bsln}$.

Then, we start again from the intrinsic-stochasticity-averaged firing rate of any recurrent neuron i_{p_o}, in response to a fixed input history $\left\{h_{i_{p_o},\,ss}(t')\right\}_{\forall t' \leq t}$ (see Equation 8.37): $r_{(i_{p_o},ss)\,|\,\left\{h_{i_{p_o},\,ss}(t')\right\}_{\forall t' \leq t}} := E_{S_{i_{p_o},\,ss}\,|\,\left\{h_{i_{p_o},\,ss}(t')\right\}_{\forall t' \leq t}}\left[S_{i_{p_o},\,ss}(t)\right]$.

We further remind the reader that:

$$E_{pop\ nrn\ p_o,\,ss} := E_{\{h_{p_o}(t')\}_{\forall t' \leq t},\,ss}\left[E_{S_{i_{p_o},\,ss}\,|\,\left\{h_{i_{p_o},\,ss}(t')\right\}_{\forall t' \leq t}}\left[\cdot\right]\right] \tag{8.44}$$

Hence, for any recurrent population p_o, we can write, using Equation 8.36 and Equation 8.37:

$$
\begin{aligned}
&E_{pop\ nrn\ p_o,\ ss}\left[r_{(i_{p_o},ss)\ |\ \left\{h_{i_{p_o},\ ss}(t')\right\}_{\forall t'\leq t}}\right] \approx \\
&E_{pop\ nrn\ p_o,\ ss}\left[\left(\Lambda_{p_o} * h_{i_{p_o},\ ss}\right)(t) + \left(R_{p_o,\ bsln} + \left(\Lambda_{p_o} * \epsilon\left(\Delta Exc_{\ i_{p_o},\ ss}\right)\right) - h_{p_o,\ bsln}\ \int \Lambda_{p_o}\right)\right] \approx \\
&\left(\Lambda_{p_o} * E_{pop\ nrn\ p_o}\left[h_{i_{p_o},\ ss}\right]\right) + \left(R_{p_o,\ bsln} + \frac{A_{p_o,\ fluct}}{\int \breve{\eta}_{p_o}} - h_{p_o,\ bsln}\ \int \Lambda_{p_o}\right) \approx \\
&\left(\Lambda_{p_o} * \left(\sum_{p=1}^{N_{pop}} n_{p,\ p_o}\ \ R_{p,\ ss}\ \left(\int F^{tot}_{p,\ p_o}\right)\right)\right) + \left(R_{p_o,\ bsln} + \frac{A_{p_o,\ fluct}}{\int \breve{\eta}_{p_o}} - h_{p_o,\ bsln}\ \int \Lambda_{p_o}\right) \approx \\
&\left(\sum_{p=1}^{N_{pop}} n_{p,\ p_o}\ \ R_{p,\ ss}\ \left(\int F^{tot}_{p,\ p_o}\right)\right)\left(\int \Lambda_{p_o}\right) + \left(R_{p_o,\ bsln} + \frac{A_{p_o,\ fluct}}{\int \breve{\eta}_{p_o}} - h_{p_o,\ bsln}\ \int \Lambda_{p_o}\right) \approx R_{p_o,\ ss}
\end{aligned}
\tag{8.45}
$$

which approximately gives the $A_{p,\ fluct}$ as a linear function of the $R_{p,\ ss}$. This permits a replacement within the C_p in Equation 8.43. Indeed:

$$
A_{p_o,\ fluct} \approx \left(\int \breve{\eta}_{p_o}\right)\left(\left(R_{p_o,\ ss} - R_{p_o,\ bsln}\right) + \left(\int \Lambda_{p_o}\right)\left(h_{p_o,\ bsln} - \sum_{p=1}^{N_{pop}} n_{p,\ p_o}\ \ R_{p,\ ss}\ \left(\int F^{tot}_{p,\ p_o}\right)\right)\right)
\tag{8.46}
$$

Hence, we can compute the $R_{p,\ ss}$ through solving Equation 8.43 while the $A_{p,\ fluct}$ are replaced with the right-hand side of Equation 8.46. Finally, one can deduce back the approximate $A_{p,\ fluct}$ through Equation 8.46.

This is necessary because the $C_p := A_{p,\ fluct} + \left(\int \breve{\eta}_p\right)\ \left(R_{p,\ bsln} - h_{p,\ bsln}\ \int \Lambda_p\right)$ are still undetermined, as $A_{p,\ fluct}$ depends on the amplitude of the fluctuations (see Equation 8.36). Indeed, $A_{p,\ fluct}$ is an estimation of the average error for the linear estimation of the adaptation's time-course (which accounts for the first-order of a Taylor expansion, see Equation 8.31 and above). This error is averaged over the single neuron's dynamic firing rates while the population is in a steady-state characterized by a variability of the received synaptic input between neurons and over time. In our case where all neurons are statistically identical, the time and the population variabilities should actually have the same properties. Hence, $A_{p,\ fluct}$ is an approximation for both the neuron-averaged, and the time-averaged, error for the estimated adaptation through the first-order of a Taylor expansion.

We note that even though this self-consistent computation of $A_{p,\ fluct}$ mitigates the error made when Taylor-expanding the time-dependent adaptation variable

to first order (in Equation 8.38), an underestimation of the absolute magnitude of this adaptation variable (and therefore a probable overestimation of the firing rate) still subsists. Indeed, we use $A_{p,\,fluct}$ as a proxy for $\breve{\eta}_p * (\Lambda_p * \epsilon\,(\Delta Exc_{\,i_p}))$ in Equation 8.38 and therefore in Equation 8.39. This is similar to enforcing, when estimating the absolute magnitude of the adaptation effect:

$$E_{pop\ nrn\ p,\,ss}\left[\exp\left(error\left(neuron\right)\right)\right] \approx \exp\left(E_{pop\ nrn\ p,\,ss}\left[error\left(neuron\right)\right]\right) \quad (8.47)$$

where $error\left(neuron\right)$ is a neuron-dependent positive error term for the absolute magnitude of adaptation which results from the systematic underestimation of the single-neuron firing rate. This underestimation comes itself from the approximation of the single-neuron firing rate by a first-order Taylor expansion of an exponential function (see the text centered on Equation 8.31). Then, we note that the (positive) derivative of the exponential function increases for larger arguments. Hence, for any distribution (over different neurons of the population p) of the absolute error that does not show a very fat tail for lower values, the left-hand-side of Equation 8.47 is larger than its right-hand-side. This means that we are underestimating $E_{pop\ nrn\ p,\,ss}\left[\exp\left(error\left(neuron\right)\right)\right]$, which then leads to an underestimation of the magnitude of adaptation effect and an overestimation of the population firing rate.

Reduction of the mathematical expressions for the dynamic rate to differential equations

Our main finding from the previous section is that the time-dependent firing rate for the recurrent populations can be written as:

$$\begin{aligned}
&R_{p_o}\left(t\right) \approx \lambda_{0,\,p_o}\ \exp\left(C_{p_o}\right)\ \exp\left(\sum_{p=1}^{N_{pop}} n_{p,\,p_o}\ \ \left(\Phi_{p,\,p_o} * R_p\right) + \frac{\sum_{p=1}^{N_{pop}} n_{p,\,p_o}\Phi\Phi_{p,\,p_o} * R_p}{2}\right) \Leftrightarrow \\
&R_{p_o}\left(t\right) \approx \lambda_{Eff,\,p_o}\ \exp\left(\sum_{p=1}^{N_{pop}} n_{p,\,p_o}\left(\Phi_{p,\,p_o} + \frac{\Phi\Phi_{p,\,p_o}}{2}\right) * R_p\right) \Leftrightarrow \\
&R_{p_o}\left(t\right) \approx \lambda_{Eff,\,p_o}\ \exp\left(I_{p_o}^{filt}(t) + \sum_{p_{recc}=1}^{N_{pop\ recc}} n_{p_{recc},\,p_o}\left(\Phi_{p_{recc},\,p_o} + \frac{\Phi\Phi_{p_{recc},\,p_o}}{2}\right) * R_{p_{recc}}\right)
\end{aligned} \quad (8.48)$$

where Φ and $\Phi\Phi$ are filters that account for the effect of the mean and the variance of the synaptic input in the population, respectively, and $I_{filt}(t)$ regroups the sum of the filtered contributions from the external populations.

We remark that as long as these filters Φ and $\Phi\Phi$ can be approximated by sums of exponential filters, the mean and variance of each Z variable can both be expressed as a sum of the solutions of differential equations where the R_p are additive variables in the derivatives ([Toyoizumi et al. (2009)]). This approximation may be made numerically, but the relative simplicity of our expressions might also allow us to link the major time-scales of these kernels to the single neuron model parameters in the future. To this aim, we would need to rework our kernel $\lambda_{0,\,p_o}$, probably in the Fourier domain, to approximately match it to the mathematical expression of the Fourier transform of an exponential kernel.

We will actually show the shape of some kernels in the results (in Figure 9.4); qualitatively speaking, an exponential basis for these kernels appears reasonable, and a good approximation could probably be reached through a few number of exponential kernel.

Indeed, to illustrate how the reduction to non-linear differential equations arises, we take a case when, for any populations p and p_o:

$$\begin{aligned} n_{p,\,p_o}\left(\Phi_{p,\,p_o}+\frac{\Phi\Phi_{p,\,p_o}}{2}\right)(s) &\approx \sum_{k\in N_{p,\,p_o}}\left(C_k^{p,\,p_o}\exp(-\frac{s}{\tau_k^{p,\,p_o}})\right)\,\Theta(s) \\ n_{p,\,p_o}\left(\Phi_{p,\,p_o}+\frac{\Phi\Phi_{p,\,p_o}}{2}\right)(s) &\approx \sum_{k\in N_{p,\,p_o}} E_k^{p,\,p_o}(s) \end{aligned} \tag{8.49}$$

where Θ is the Heaviside function, $N_{p,\,p_o}$ is the number of exponentials needed to fit $\left(\Phi_{p,\,p_o}+\frac{\Phi\Phi_{p,\,p_o}}{2}\right)$, C and τ are constant (with $\tau>0$), and $E_k^{p,\,p_o}(s) := \left(C_k^{p,\,p_o}\exp(-\frac{s}{\tau_k^{p,\,p_o}})\right)\,\Theta(s)$.

Then, we can write, for any populations p and p_o, and any $k\in N_{p,\,p_o}$:

$$\begin{aligned} V_k^{p,\,p_o} &:= E_k^{p,\,p_o} * R_p \Rightarrow \\ \frac{dV_k^{p,\,p_o}}{dt} &= -\frac{V_k^{p,\,p_o}}{\tau_k^{p,\,p_o}} + C_k^{p,\,p_o} R_p(t) \end{aligned} \tag{8.50}$$

which can be verified by simple differentiation of $V_k^{p,\,p_o}$.

Finally, using Equation 8.48 and Equation 8.49, we can rewrite any the

recurrent populations p_1 and p_o the derivative of $V_k^{p_{recc},\,p_o}$:

$$\frac{dV_k^{p_1,\,p_o}}{dt} \approx -\frac{V_k^{p,1\ p_o}}{\tau_k^{p_1,\,p_o}} + C_k^{p_1,\,p_o}\lambda_{Eff,\,p_1}\ \exp\left(I_{p_1}^{filt}(t) + \sum_{p_{recc}=1}^{N_{pop\ recc}} V_k^{p_{recc},\,p_1}(t)\right) \tag{8.51}$$

with, for each recurrent population p_o and at each time, $R_{p_o}(t) \approx \lambda_{Eff,\,p_o}\ \exp\left(I_{p_o}^{filt}(t) + \sum_{p_{recc}=1}^{N_{pop\ recc}} V_k^{p_{recc},\,p_o}(t)\right)$.

Hence, the different V variables define a system of coupled non-linear differential equation that may be studied with the usual stability analysis tools (linear stability, phase plane).

Finally, these equations may be linearized through a Taylor-expansion of the exponential if needed. Note that, in contrast to Equation 8.37, this linearization would retain a contribution from the variance of the synaptic input.

8.3 Comparison between analytics and network simulations

In order to test our formulas, we decided to use a rather simple network simulation paradigm, to facilitate the comparison with our equations. Hence, we used a simple recurrent network connected with inhibitory synapses, that receive Poisson spike trains from external sources connecting to both excitatory and inhibitory synapses. This type of network can easily maintain asynchronous irregular activity [Brunel and Hakim (1999)], which is one of the requirements for the validity of our approximations. Due to time constraints, all ranges of validity could not be extensively tested yet. We describe here the choices we made.

We will start by describing the parameters for the internal dynamics of the neuron model which was used for the recurrent population. We then turn to describe the connectivity of the network. Finally, we explain how we chose some specific types of temporal modulations for the firing rate of the external populations.

8.3.1 Internal dynamics' parameters for the single neuron

We hereby describe the parameters taken for the recurrently connected Generalized Linear Model adapting neurons (see Equation 8.1).

Shape and amplitude of the spike-history filter

We chose a "power-law like" spike-history filter which shape and amplitude were approximately matched to those of 10 recorded pyramidal neurons (data courtesy of S. Mensi), but we chose to keep "only" three time-scales for this kernel for reducing computation time. Note that the stimulations used for fitting made the neurons fire at a relatively low steady-state firing rate ($\approx$ 10 Hz and smaller, but the short-term modulations can be much larger, see subsection 8.1.3, Mensi et al. (2012)). Note that the parameters we use are not 'round' as a result of an initial attempt to take the two first exponentials as the best fit to the kernel extracted from a recorded pyramidal neuron (we have not been trying to optimize the match with the theory...). The amplitude for the slowest exponential was chosen to be slightly larger than what is generally observed in pyramidal neurons (see Figure 8.2), in order to still keep a small negative amplitude for very long delays after a spike (which could also be implemented by adding more exponential variables to the filter). The aim was not to have a perfect quantitative match to the recorded neurons, but rather to see whether a realistic spike-history filter would still make the simplifications that we used for deriving the approximate mean-field formulas acceptable.

More precisely, the spike history filter η, drawn in Figure 8.2 was taken as (see Equation 8.1 for the definition of the model):

$$\eta(s) = \left(-33.55\exp\left(-\frac{s}{4.9\ ms}\right) - 5\exp\left(-\frac{s}{60\ ms}\right) - 2\exp\left(-\frac{s}{300\ ms}\right)\right)\Theta(s) \tag{8.52}$$

where $\Theta(s)$ is the Heaviside function.

Shape and amplitude of the combined leak-and-synapse filter

By definition, the combined leak-and-synapse filters $F^{tot}_{p,\,p_o}$ from one neuron of population p to one neuron of a population p_o are the result of a convolution between the membrane and synaptic filters scaled by the intrinsic noise of the

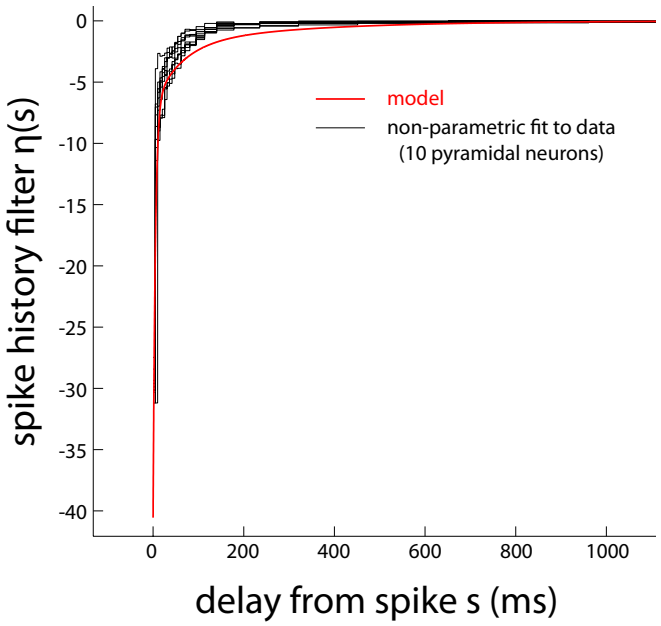


Figure 8.2: *Comparison of the spike history kernel used in the simulations.* Here, we show in red the spike-history kernel η (see Equation 8.52 and Equation 8.1) that we used in the simulation. For comparison, we show the values of η fitted non-parametrically in 10 pyramidal neurons (data courtesy from S. Mensi).

neuron (see Equation 8.9). Hence, $F^{tot}_{p,\,p_o}$ corresponds to the time-course of a Post Synaptic Potential (PSP) measured of the soma, and its amplitude actually corresponds to the amplitude of one PSP divided by the intrinsic noise ΔV (see Equation 8.4). Hence, $F^{tot}_{p,\,p_o} := \left(\frac{\kappa}{\Delta V} * F_{p,\,p_o}\right)$. Here, ΔV is the intrinsic noise, and $F_{p,\,p_o}$ is the time-course of the effective current received at the soma that results from synaptic channels opening. Finally, $\kappa\left(s\right) := \frac{\exp\left(-\frac{s}{\tau}\right)}{C}\,\Theta\left(s\right)$ is the membrane filter, where C is the conductance of the neuron, $\tau = \frac{C}{g_L}$ is the membrane time-scale while g_L is the leak conductance, and Θ is the Heaviside function. Indeed, one can verify that when taking $V_{subthld} = \kappa * I\left(t\right)$, where I is an input current, then $C\,\frac{dV_{subthld}}{dt} = -g_L\,V_{subthld} + I\left(t\right)$.

It is customary to take an exponential shape for both the synaptic and the membrane filters, which implies to write $F_{p,\,p_o}\left(s\right) := A_{syn}\ \exp\left(-\frac{s}{\tau_s}\right)\ \Theta\left(s\right)$. Then, the combined filter $F^{tot}_{p,\,p_o}$ is a difference of the two exponentials, with the final decay occurring with the largest time-scale. Indeed, using the definition of the membrane filter and performing a convolution with the synaptic filter (for $\tau \neq \tau_s$), we get:

$$
\begin{aligned}
F^{tot}_{p,\,p_o}(t) &:= \int_0^t \left(\frac{1}{C\,\Delta V}\,\exp\left(-\frac{s}{\tau}\right)\,A_{syn}\,\exp\left(-\frac{t-s}{\tau_s}\right)\right) ds \\
F^{tot}_{p,\,p_o}(t) &= \frac{A_{syn}}{C\,\Delta V}\,\exp\left(-\frac{t}{\tau_s}\right)\int_0^t \exp\left(-s\left(\frac{1}{\tau}-\frac{1}{\tau_s}\right)\right) ds \\
F^{tot}_{p,\,p_o}(t) &= \frac{A_{syn}}{C\,\Delta V}\,\exp\left(-\frac{t}{\tau_s}\right)\left(-\frac{1}{\frac{1}{\tau}-\frac{1}{\tau_s}}\right)\left(\exp\left(-t\left(\frac{1}{\tau}-\frac{1}{\tau_s}\right)\right)-1\right) \\
F^{tot}_{p,\,p_o}(t) &= \frac{A_{syn}}{C\,\Delta V}\,\frac{\tau\,\tau_s}{max(\tau,\,\tau_s)-min(\tau,\,\tau_s)}\left(\exp\left(-\frac{t}{max(\tau,\,\tau_s)}\right)-\exp\left(-\frac{t}{min(\tau,\,\tau_s)}\right)\right)
\end{aligned}
\tag{8.53}
$$

As long as there is a large enough difference between $max(\tau,\,\tau_s)$ and $min(\tau,\,\tau_s)$, this $F^{tot}_{p,\,p_o}$ will have a shape close to a single exponential (with a decay of time scale $\approx max(\tau,\,\tau_s)$, and an amplitude that can be adjusted so as to get the same integral as the original difference of exponentials). This is the approximation we made. Note that this also implies that we neglected synaptic delays (which may be implemented through setting the initial bins of the synaptic filter to 0).

We considered cases when synaptic transmission has a slow component [Wang et al. (2008, 2013)], which may be the major source of synaptic current even if it has a modest amplitude (as the integral of $F_{p,\,p_o}$ also scales with τ_s, see for instance the NMDA channel in [Brunel and Wang (2001)]). Note that in our network the long time-scale recurrent connections are inhibitory (which is different from the phenomena described in the above-mentioned publications), but we still wanted to include a slow component of the synaptic input in order to the verify that the theory could handle it.

We therefore took two relatively long time-constant for both the recurrent connections and the external excitatory connections (with respectively 70 and 50 ms). For external inhibition, we considered a case when the slowest component of $F^{tot}_{p,\,p_o}$ is dominated by the membrane time-scale, which we took to be around 20ms. In future settings, it will be necessary to test shorter time scales (for synapses without slow component, and to account for the shorter effective membrane time-scale shaped by the high-conductance state in vivo Destexhe et al. (2003)).

Finally, we took amplitudes for the $F^{tot}_{p,\,p_o}$ that were close to the amplitude of unitary PSPs as measured in awake monkeys at the soma ($\approx$ 0.1-0.3 mV). These data were obtained through simultaneous intracellular and extracellular recordings in motor cortex [Matsumura et al. (1996)], and were therefore probably

mostly reflecting unitary PSPs in pyramidal neurons. In addition, the order of magnitude for the PSP amplitude is consistent with more recent data for PSPs in inhibitory neurons in the barrel cortex of anesthetized mice ($\approx 0.4mV$, Pala and Petersen (2015)). Note that, in real neurons, the amplitude of these PSPs depends on the regime of synaptic bombardment, which is the reason why we searched for in vivo data.

$$\begin{aligned}
(\Delta V)\ \left(F^{tot}_{p_{recc},\, p_{recc}}(s)\right) &= -\,(0.1\ mV)\ \exp\left(-\frac{s}{70\ ms}\right) \\
(\Delta V)\ \left(F^{tot}_{p_{ext\ exc},\, p_{recc}}(s)\right) &= (0.11\ mV)\ \exp\left(-\frac{s}{50\ ms}\right) \\
(\Delta V)\ \left(F^{tot}_{p_{ext\ inh},\, p_{recc}}(s)\right) &= -\,(0.2\ mV)\ \exp\left(-\frac{s}{20\ ms}\right)
\end{aligned} \tag{8.54}$$

where the subscripts *recc*, *ext exc* and *ext inh* stand for the recurrent population, the external excitatory input and the external inhibitory input, respectively; and $\Delta V \approx 1$mV [Pozzorini et al. (2013)] is the intrinsic noise.

Baseline firing rate $\lambda_{0,\, p_{recc}}$

On the same ensemble of neurons as those showing the kernels of Figure 8.2, the mean λ_0 was $(\exp(-5.2)\ Hz)$. We took $(\lambda_{0,\, p_{recc}} = \exp(-5)\ Hz)$.

Note that this parameter is equivalent to a baseline mean filtered input (that would be the same in all neurons).

8.3.2 Network connectivity and number of neurons

We chose to have a fixed number of connections for each of the $n_{recc} = 2000$ recurrent neurons. Each of those was receiving $n_{p_{ext\ exc},\, p_{recc}} = 1000$ external excitatory inputs, $n_{p_{ext\ inh},\, p_{recc}} = 400$ external inhibitory inputs, and $n_{p_{recc},\, p_{recc}} = 0.3 \cdot 2000 = 600$ recurrent inhibitory inputs. Hence, we are working with rather large numbers of inputs, which could allow a convergence of the Central Limit Theorem for the synaptic input (see subsection 8.2.3), but that still represent only a few percent of the total number of inputs received by cortical pyramidal neurons [Megías et al. (2001)].

8.3.3 Design of external firing rate simulations

The external populations were modeled as independent inhomogeneous Poisson spike trains.

We first used either steady-state stimulations, where the external units were all firing at constant rates $R_{ext\ exc}$ and $R_{ext\ inh}$. We characterized such inputs by the mean and variance of the filtered external input received by each recurrent neuron $I^{ext}_{i_{recc}}$ (which is in units of the intrinsic noise ΔV, as F^{tot} is the convolution between the membrane and synaptic filters divided by ΔV):

$$I^{ext}_{i_{recc}}(t) := \sum_{j=1}^{n_{p_{ext\ exc}},\, p_{recc}} F^{tot}_{p_{ext\ exc},\, p_{recc}} * S_j^{p_{ext\ exc},\, i_{recc}} + \sum_{j=1}^{n_{p_{ext\ inh}},\, p_{recc}} F^{tot}_{p_{ext\ inh},\, p_{recc}} * S_j^{p_{ext\ inh},\, i_{recc}} \tag{8.55}$$

From this definition, we can calculate the mean and the variance of the intrinsic-noise-scaled voltage caused by external inputs:

$$E\left[I^{ext}_{i_{recc}}\right] = n_{p_{ext\ exc},\, p_{recc}} \left(\int F^{tot}_{p_{ext\ exc},\, p_{recc}}\right) R_{ext\ exc} + n_{p_{ext\ inh},\, p_{recc}} \left(\int F^{tot}_{p_{ext\ inh},\, p_{recc}}\right) R_{ext\ inh}$$

$$var\left[I^{ext}_{i_{recc}}\right] = n_{p_{ext\ exc},\, p_{recc}} \left(\int FF^{tot}_{p_{ext\ exc},\, p_{recc}}\right) R_{ext\ exc} + n_{p_{ext\ inh},\, p_{recc}} \left(\int FF^{tot}_{p_{ext\ inh},\, p_{recc}}\right) R_{ext\ inh} \quad where\ \forall s,\ FF^{tot}(s) := \left(F^{tot}\right)^2 \tag{8.56}$$

Note that, in steady-state, and only in steady-state, the means and variances over neurons are the same as the mean and variances over time in our network where all neurons are statistically identical.

We also used time-dependent stimulations, with firing rates that were modulated according to either an Ornstein-Uhlenbeck process, or a simple sine wave. Given a time-dependent excitatory rate $R_{ext\ exc}(t)$, we were particularly interested in creating a stimulus for which the dynamics would be driven by a variance of the driving current $I^{ext}_{i_{recc}}(t)$. We were then interested in finding $R_{ext\ inh}(t)$ such that $E\left[I^{ext}_{i_{recc}}(t)\right] = \alpha$ is a constant (i.e., it does not depend on

time). We solved this problem in the Fourier space, where it becomes simpler:

$$E\left[I_{i_{recc}}^{ext}(t)\right] = \alpha \;\Rightarrow$$

$$n_{p_{ext\ exc},\, p_{recc}} \mathfrak{F}\left[F^{tot}_{p_{ext\ exc},\, p_{recc}} * R_{ext\ exc}\right] + n_{p_{ext\ inh},\, p_{recc}} \mathfrak{F}\left[F^{tot}_{p_{ext\ inh},\, p_{recc}} * R_{ext\ inh}\right] = \mathfrak{F}[\alpha] \;\Leftrightarrow$$

$$n_{p_{ext\ exc},\, p_{recc}} \mathfrak{F}\left[F^{tot}_{p_{ext\ exc},\, p_{recc}}\right] \mathfrak{F}[R_{ext\ exc}] + n_{p_{ext\ inh},\, p_{recc}} \mathfrak{F}\left[F^{tot}_{p_{ext\ inh},\, p_{recc}}\right] \mathfrak{F}[R_{ext\ inh}] = \mathfrak{F}[\alpha] \;\Leftrightarrow \tag{8.57}$$

$$\mathfrak{F}[R_{ext\ inh}] = \frac{\mathfrak{F}[\alpha] - n_{p_{ext\ exc},\, p_{recc}} \mathfrak{F}\left[F^{tot}_{p_{ext\ exc},\, p_{recc}}\right] \mathfrak{F}[R_{ext\ exc}]}{n_{p_{ext\ inh},\, p_{recc}} \mathfrak{F}\left[F^{tot}_{p_{ext\ inh},\, p_{recc}}\right]} \;\Leftrightarrow$$

$$R_{ext\ inh}(t) = \mathfrak{F}^{-1}\left[\frac{\mathfrak{F}[\alpha] - n_{p_{ext\ exc},\, p_{recc}} \mathfrak{F}\left[F^{tot}_{p_{ext\ exc},\, p_{recc}}\right] \mathfrak{F}[R_{ext\ exc}]}{n_{p_{ext\ inh},\, p_{recc}} \mathfrak{F}\left[F^{tot}_{p_{ext\ inh},\, p_{recc}}\right]}\right]$$

Note that one has to be careful to get physically meaningful results for the inhibitory firing rate, i.e. the rates should be positive. This requires to impose a reachable value of α (which should be at any time smaller than the population-averaged excitatory filtered input).

8.3.4 Numerics

Neuronal network simulation

We used the Brian neural simulator, version 2 beta 2 (http://brian2.readthedocs.org/en/latest/introduction/index.html). The kernels were implemented as sums of exponentials, which hence correspond to the solutions of linear differential equations (see section 8.3.1 for an example). These equations were integrated numerically with a time-step $dt = 0.1$ ms. the performance was optimized by using the scipy.weave code package to run C code instead of native python code.

Convolutions, fourier transformation and power spectral density

Numerical operations were performed with python packages.

We used the same dt (10^{-4} s) for the numerical operations as for the network simulation. Convolutions were performed with the numpy function convolve. Fourier transformation used the numpy.fft function rfft (for real numbers). For the power spectral density, we used the function psd from the matplotlib.pylab library, which uses the Welchs average periodogram method. We used a block size for fast fourier transform computation that was a multiple of the stimulation period that we had imposed, and that corresponded to about 20% of the total number of data points. The overlap between blocks was 25%.

Chapter 9

Tests and applications of the new analysis tools for adapting neuronal network dynamics

---

We developed approximate analytical expressions for the population firing rate in recurrent networks of adapting spiking Generalized Linear Model neurons. For each neuron i_{recc} of one recurrent population, the input spikes arriving at the synapses are filtered through a combined leak-and-synapse filter F^{tot}, and then summed, leading to a total synaptic drive $h_{i_{recc}}$. This filtered synaptic input is added to an adaptation variable wich results from the filtering of the spike train of the neuron $S_{i_{recc}}$ through a spike-history filter η_{recc}. Finally, the probability of spiking is proportional to $\exp\left(h_{i_{recc}} + \eta_{recc} * S_{i_{recc}}\right)$ (see Equation 8.3 for details).

To sum up the mathematical methods described in chapter 8, we consider the convergence of the filtered synaptic input to a Gaussian distribution (which is valid in case of a large number of synapses, see subsection 8.2.3). In addition, we use a number of approximations for the time-course of the adaptation variable in order to reach non-linear population firing rate equations. One strength of the approach is that adaptation is not considered as stationary, and we can account for the effects of the variability of the synaptic input within the populations of neurons. In addition, we can predict and understand when and how the simulations will diverge from our mathematical expressions. We summarize here the approximations that we made. These approximations would lead to an inaccuracy of the predicted upcoming firing rate even if we were able to use an exact value of past recurrent firing rates for the computation.

1. We use the first moment only of a moment-based expansion for the adaptation variable averaged over the intrinsic stochasticity of a single neuron (see Equation 8.22). As discussed in the Methods, given the negative spike-time correlations expected with adaptive neurons, this

approximation is likely to lead to an underestimation of the adaptation term and therefore to an overestimation of the population firing rate (e.g. see Figure 8.1).

2. We linearize the averaged adaptation (over the intrinsic stochasticity of a single neuron, see Equation 8.38, Equation 8.46 and Equation 8.47). This is likely to lead to a slight underestimation of the adaptation, which in turn should lead to an overestimation of the population firing rate. This will be the case as long as the voltage deviations relative to baseline are not overly biased towards negative values.

3. We assume that the spiking is approximately inhomogeneous Poisson (see Equation 8.11). Given that adaptation is likely to produce negative correlations between spike times, we will tend to overestimate the variance of the synaptic input. This overestimation of the variance would also tend to create an overestimation of the firing rates (see Equation 8.42).

Hence, all approximations consistently lead to an overestimation of the firing rate. This may either be seen as a curse, or as an advantage. Indeed, on one hand, the different deviations will sum up and cannot compensate for one another. On the other hand, the effect of these deviations is predictible: an overestimated firing rate. In addition, the amount of this overestimation is expected to grow with the strength of adaptation and with the firing rate. Interestingly, while this is likely to affect the prediction of the amplitude of the firing rate, the approximations are unlikely to affect much the time-course of the firing rate modulations. In other words, the sign of the time-derivative of the firing rates is expected to be well-preserved. In contrast, if the different approximations would have had antagonistic effects with different amplitudes at different times, the shape of the firing rate time-course could have been expected to be distorted. For instance, there could have been an overall overestimation of the rate at the beginning of a stimulation, later followed by underestimation of the firing rate.

We would like to mention that we used a very strong spike-history kernel η to implement the adaptation, even compared to previous publications ($\int \eta$ is 1.66 times the one used in Naud and Gerstner (2012a), see Figure 8.2). An example of the dynamics of such a neuron model embedded in a network with a single recurrent inhibitory population settling in an asynchronous irregular state can be found in Figure 9.1. The different neurons indeed appear to fire at different times (Figure 9.1 (c)), and fire irregularly (Figure 9.1 (d)). Note that, throughout

the Results section, we will use the subscript "*recc*" to mark the variables of the (single) recurrent population. Hence, the more general subscript p_o used in the Methods section can now be replaced by "*recc*". In addition, the subscripts "*ext exc*" and "*ext inh*" will be used to refer to the external excitatory and inhibitory input populations.

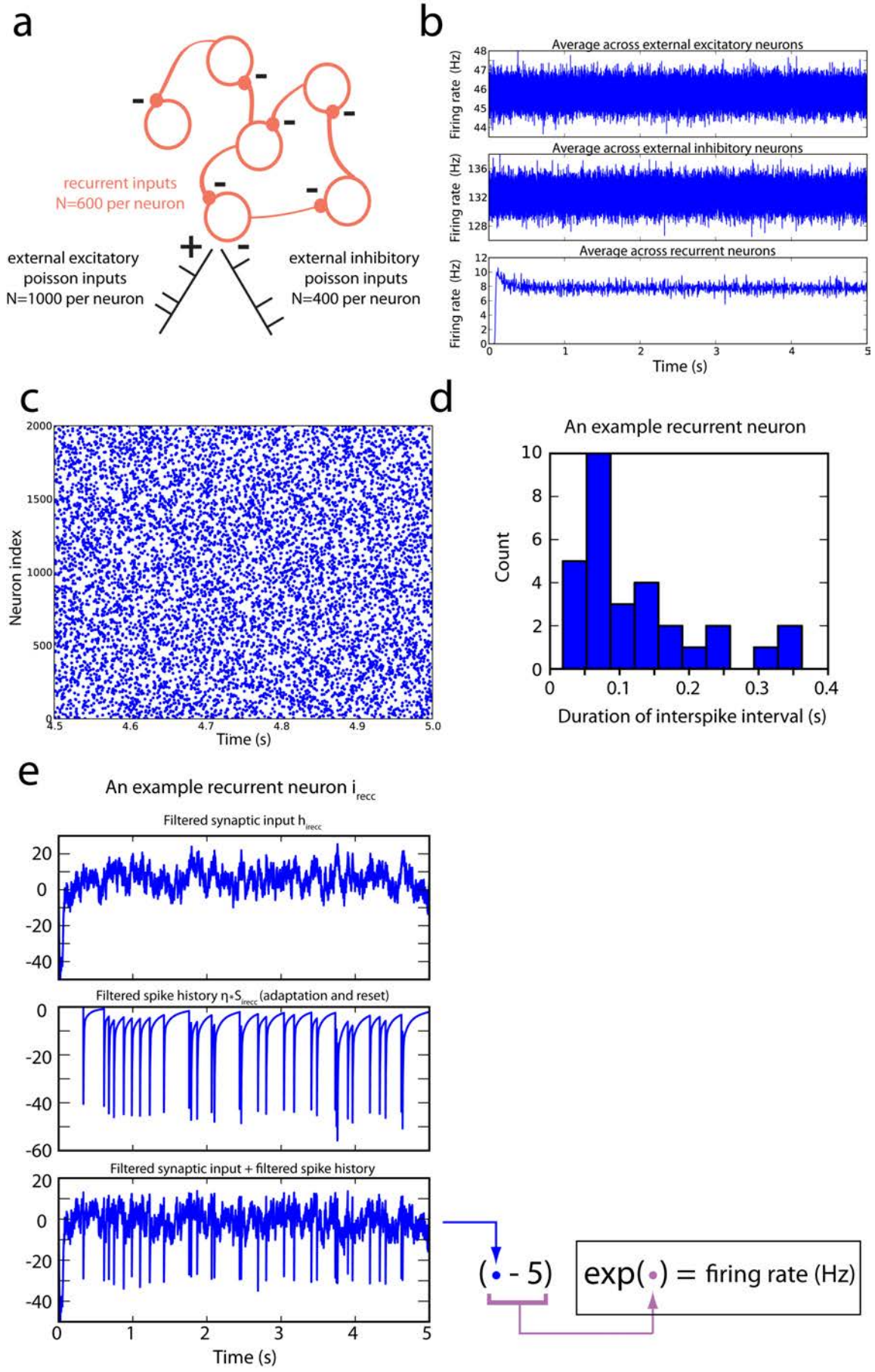

a
recurrent inputs
N=600 per neuron
external excitatory
poisson inputs
N=1000 per neuron
external inhibitory
poisson inputs
N=400 per neuron
b
Average across external excitatory neurons
Firing rate (Hz)
Average across external inhibitory neurons
Firing rate (Hz)
Average across recurrent neurons
Firing rate (Hz)
Time (s)
c
Neuron index
Time (s)
d
An example recurrent neuron
Count
Duration of interspike interval (s)
e
An example recurrent neuron i recc
Filtered synaptic input h irecc
Filtered spike history η*S irecc (adaptation and reset)
Filtered synaptic input + filtered spike history
Time (s)
(• - 5)
exp(•) = firing rate (Hz)

Figure 9.1 *(previous page)*: *An example simulation of a network of Generalized Linear Model neurons with adaptation.* **(a)** Network architecture. 2000 GLM neurons are recurrently connected by inhibitory synapses. Each of these neurons receives 600 recurrent connections, as well as 1000 excitatory and 400 inhibitory external inputs. The external inputs are uncorrelated between neurons, and are modeled as Poisson processes (which are constant over time in this particular figure, but that can be time-dependent in general). **(b)** Average firing rate in the three types of neurons (external excitatory, external inhibitory and recurrent). For clarity, the recurrent rates were estimated in bins of length 5 ms. **(c)** Raster plot for the network activity. Each line corresponds to a recurrent neuron, a dot indicates that the neuron has fired at the time of the abscissa. **(d)** An example of interspike interval distribution from one recurrent neuron (taken in the steady-state, over the last 3.8 s of the simulation). This shows that the interspike intervals are variable, and confirms the irregular firing behavior. **(e)** The internal variables governing the spiking of one recurrent neuron i_{recc}. Top: filtered synaptic input $h_{i_{recc}}$ (see Equation 8.16; this includes both the external and the recurrent synaptic input). The filtering occured through the combined leak-and-synapse filter F^{tot} (each population is associated with a specific leak-and-synapse filter). Despite the constant external firing rate modulation, the filtered input varied a lot over time due to the presence of both excitatory and inhibitory inputs. More precisely, once the steady-state was reached, the F^{tot}-filtered external input had a mean of 40 and a variance (over neurons) of 35; note the sizable reduction of this mean in $h_{i_{recc}}$ through the addition of the recurrent filtered input. Middle: filtered spike history $\eta * S_{i_{recc}}$ (see Equation 8.1). There is a clear cumulative effect over several spikes, which is a hallmark of adaptation. Indeed, this differs from refractoriness properties, which can be modeled as a function of the last spike time only. Bottom: sum of the top and middle signals. At the right of this plot, we illustrate that the instantaneous firing rate, in hertz, was the exponential of the sum of two components: *(i)* the variable displayed in the bottom graph and *(ii)* a constant threshold value of -5. Finally, at each time-step, the probability of firing was the instantaneous firing rate times the time-step duration in seconds ($10^{-4}s$ in our case, see Equation 8.1 and subsection 8.3.1).

Here, we test how well our equations can describe the dynamics of a recurrent inhibitory population, and we present some insights reached thanks to our analysis. Note that due to constraints on the duration of the doctoral studies, the tests and results presented here are not as complete and diverse as we initially intended. We use a simple network with only one recurrently connected population of inhibitory neurons (see section 8.3 for the details). We start by testing the convergence of the filtered synaptic input to a gaussian variable. We then look at the prediction of the steady-state population firing rate. Then, we test some dynamical stimulations. Finally, we outline a few interpretations and applications allowed by our mathematical analysis.

9.1 Distribution of the sum of filtered inputs

As discussed in subsection 8.2.3, by using the Central Limit Theorem, we concluded that the summed filtered synaptic input $h_{i_{po}}(t)$ (defined in Equation 8.16) was likely to converge to a Gaussian distribution over different

neurons of the network. In addition, in case of a dynamic stimulation of the neuronal network, the parameters of this Gaussian were predicted to be time-dependent.

We first investigated the convergence of the Central Limit Theorem in a stationary firing situation, by examinating the distribution of $h_{i_{recc}}$ on the last time point of the stimulation shown in Figure 9.1.

9.1.1 Distribution of the sum of filtered inputs in a stationary regime

In Figure 9.2 (a,b,c,d), we compare the distributions (among the population of 2000 recurrent neurons) of different components of the filtered synaptic input $h_{i_{recc}}$ to Gaussian variables. These distributions were extracted from the last time-step of the simulation shown in Figure 9.1.

We first considered the shape of a Gaussian cumulative distribution function (cdf) which mean and variance would be identical to the mean and variance estimated from the simulation. The correspondence between this Gaussian variable and the distribution of inputs oberved in the simulation was almost perfect: their cdf curves were almost identical (see blue and green curves in Figure 9.2 (a,b,c,d)). This indicated that, in our simulation, there was an excellent convergence of the filtered inputs to a Gaussian distribution, in agreement with the asymptotic result of the Central Limit Theorem.

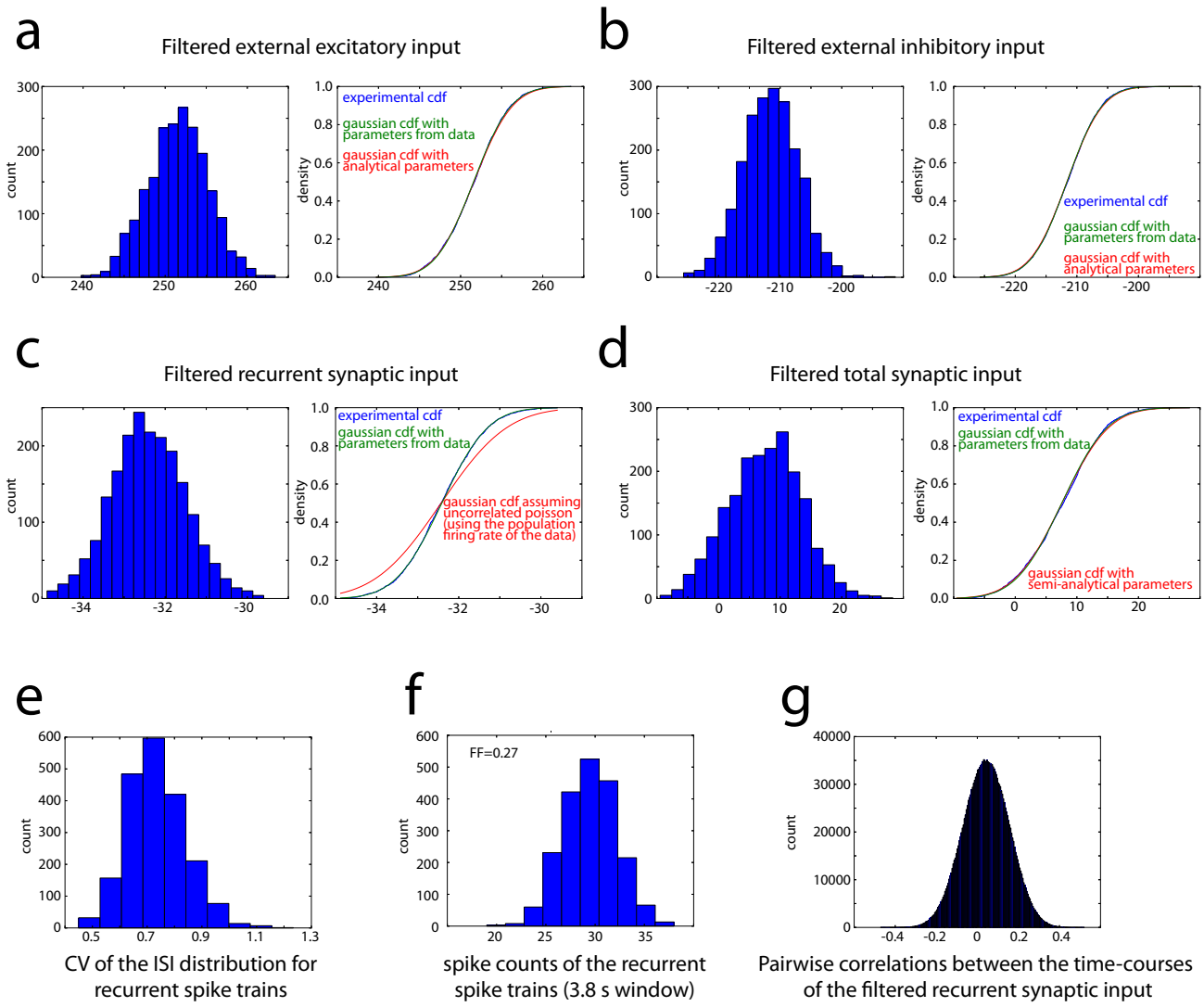


Figure 9.2: *Investigation of the shape of the distribution of filtered input in a steady-state regime.* **(a,b,c,d)** Histograms (*left*) and cumulative distribution function (cdf, *right*) for the filtered input on the last (5 s) time-point of Figure 9.1, among the 2000 recurrent neurons. We show separately:

- the filtered external excitatory input $\sum_{j=1}^{n_{ext\ exc,\ recc}} F^{tot}_{ext\ exc,\ recc} * S_j^{ext\ exc,\ i_{recc}}$ in **(a)**
- the filtered external inhibitory input $\sum_{j=1}^{n_{ext\ inh,\ recc}} F^{tot}_{ext\ inh,\ recc} * S_j^{ext\ inh,\ i_{recc}}$ in **(b)**
- the filtered recurrent synaptic input $\sum_{j=1}^{n_{recc,\ recc}} F^{tot}_{recc,\ recc} * S_j^{recc,\ i_{recc}}$ in **(c)**
- the total filtered synaptic input $h_{i_{recc}}$ in **(d)** (which is the sum of the three above-mentioned variables, see Equation 8.16)

Note that F^{tot} is a combined leak-and-synapse filter which approximates the shape of a post-synaptic potential measured at the soma as a result of the reception of a single presynaptic spike. On the cdf plot, we show both the experimental cdf (in blue), the cdf of a gaussian which mean and variance are estimated from the data (in green), and the cdf of a gaussian which mean and variance are derived from the formula for uncorrelated Poisson processes (in red, see Equation 8.17 and Equation 8.18). Note that for this computation, we used the known theoretical expected rates for the external inputs (which are those we imposed in the parameters of the simulation). In contrast, for the recurrent inputs, we used an estimation of the population-averaged rate coming from the simulation itself (we measured the recurrent population rate in the last 70 ms of the simulation of Figure 9.1). Hence, we are only examining here the accuracy of the convergence to a gaussian variable and of the uncorrelated Poisson firing approximation. Note that the three cdf curves are almost perfectly superimposed, with exception of the recurrent input, for which the Poisson approximation appears to lead to an overestimation of the variance. **(e)** Distribution of the coefficient of variation (CV) of the interspike interval distributions of single recurrent neurons (over the last 3.8 seconds of the simulation in Figure 9.1, so in steady state). The spike trains were on average more regular than a homogeneous Poisson process (which has a CV of one). **(f)** Distribution of the spike count for the recurrent spike trains (during the 3.8 last seconds of the simulation in Figure 9.1). The fano factor (FF) of this distribution is indicated; it is well below the value of 1 for a Poisson process. **(g)** Distribution of the pairwise correlation coefficients between the time-courses of the filtered recurrent synaptic input (over the last second of the simulation in Figure 9.1). All $(2000 * (2000 - 1))$ possible pairs of neurons were considered. The distribution has only a very slight bias towards positive values.

We also implemented the formulas for computing the mean and variance of the distribution assuming uncorrelated Poisson spike trains (see Equation 8.17 and Equation 8.18). As the external inputs were implemented as uncorrelated Poisson processes, these formulas should be exact in this case. Indeed, the cumulative distribution function of a Gaussian variable with these analytical parameters almost perfectly fitted the data. Indeed, the red curve was almost perfectly superimposed on the blue curve in Figure 9.2 (a,b). We also investigated the accuracy of the estimation of the mean and variance of the filtered recurrent input through the uncorrelated Poisson spiking approximation. We used Equation 8.17 and Equation 8.18 with the recurrent population firing rate taken from the 70 last ms of the simulation (70 ms is the time-scale of the filter for the recurrent input, see section 8.3.1). This led to a slight overestimation of the variance of the distribution of filtered recurrent synaptic input (as can be seen when comparing the red and the blue curves in Figure 9.2 (c)). However, this overestimation only had a negligible effect for the estimation of the variance of the total filtered synaptic input (as can be seen when comparing the red and the blue curves in Figure 9.2 (d)). Indeed, the contribution of the external excitatory and inhibitory inputs to the total input variance was large.

We further investigated the reasons why the Poisson firing approximation did not give an exact prediction of the variance of the recurrent filtered synaptic input. We plot in Figure 9.2 (e) the distribution of the coefficient of variation of the interspike interval distribution over the last 3.8 s (hence, in steady state). This amounts to computing the ratio between the standard deviation and the mean of the distribution such as the one shown in Figure 9.1 (d), for different neurons. In steady-state, for Poisson processes, this distribution should be centered on one [Gerstner et al. (2014)]. Instead, in our simulation, it was centered around 0.75, indicating that the spike trains of the recurrent neurons were more regular than what would be expected for Poisson processes. This was likely to be caused by adaptation, which can create correlations between spike times and a subsequent reduction of the coefficient of variation of the interspike interval distribution [Schwalger and Lindner (2013)]. In addition, we found that the spike count distribution (Figure 9.2 (f)) was also less variable than what would be expected from a Poisson process. Indeed, the Fano Factor (FF, defined as the ratio of the variance over the mean) of this distribution was ≈ 0.27, against 1 for a Poisson process [Farkhooi et al. (2011)]. This is again consistent with an effect of adaptation, which can "correct" an excess of spiking at time t_0

by reduced spiking at time $t > t_0$, and conversely [Farkhooi et al. (2011)]. However, in Equation 8.18 we neglect these negative spiking covariations, and hence this leads to a slight overestimation of the variance.

In addition, in Equation 8.18, we neglect between-neuron correlations of the filtered spike trains. Because recurrent neurons share a small number of connections, some positive correlations between the firing times of different neurons may be expected. Also, these positive correlations could potentially, in general, further grow through the recurrent dynamics. This might create correlations between the synaptic inputs received by different recurrent neurons. We actually found that the distribution of the (between-neuron) correlation coefficients for the time-course of the recurrent filtered synaptic input was centered close to zero (Figure 9.2 (g)). The mean correlation was actually smaller than the fraction f of shared spike trains between the recurrent filtered input of two neurons ($f = 0.3$ on average, as $0.3 = \frac{600}{2000}$ is the recurrent connection probability). This suggested that the recurrent dynamics of the network was not tending to create positive correlations between the activity of recurrent neurons. Recurrent inhibition may have, to some extent, compensated the positive correlations due to shared inputs, hence leading to relatively uncorrelated activity (a phenomenon which might be similar to the mechanisms at stake in balanced networks [Renart et al. (2010)]).

Hence, the recurrent neurons received rather uncorrelated inputs in general (with little correlations on average between recurrent inputs, and no correlations by design between external input). These uncorrelated inputs are therefore expected to lead to spike trains which are themselves uncorrelated on average. This is consistent with the observation that recurrent neurons qualitatively appeared to fire in a relatively uncorrelated fashion (as can be seen more qualitatively in Figure 9.1 (c)). Hence, discarding the summed covariance between filtered spike trains coming from different neurons (as we do in Equation 8.18) was probably reasonable.

We now generalize these results to a non-stationary regime of firing.

9.1.2 Distribution of the sum of filtered inputs in a non-stationary regime

In Figure 9.3, we show that for all components of the filtered synaptic input $h_{i_{recc}}$, the distribution over the recurrent neuronal population resembled a Gaussian with parameters that varied over time. Indeed, in a Gaussian distribution, the $16^{th} - 84^{th}$ percentiles correspond to the mean $\pm$ standard deviation confidence interval, while deviations from Gaussianity (such as asymmetry) are likely to lead to a mismatch between the two. We did observe that, at any time, the one-standard-deviation confidence interval around the mean (black curves) matched almost perfectly the $16^{th} - 84^{th}$ percentiles of the distribution (green curves), suggesting a Gaussian shape of the instantaneous distributions.

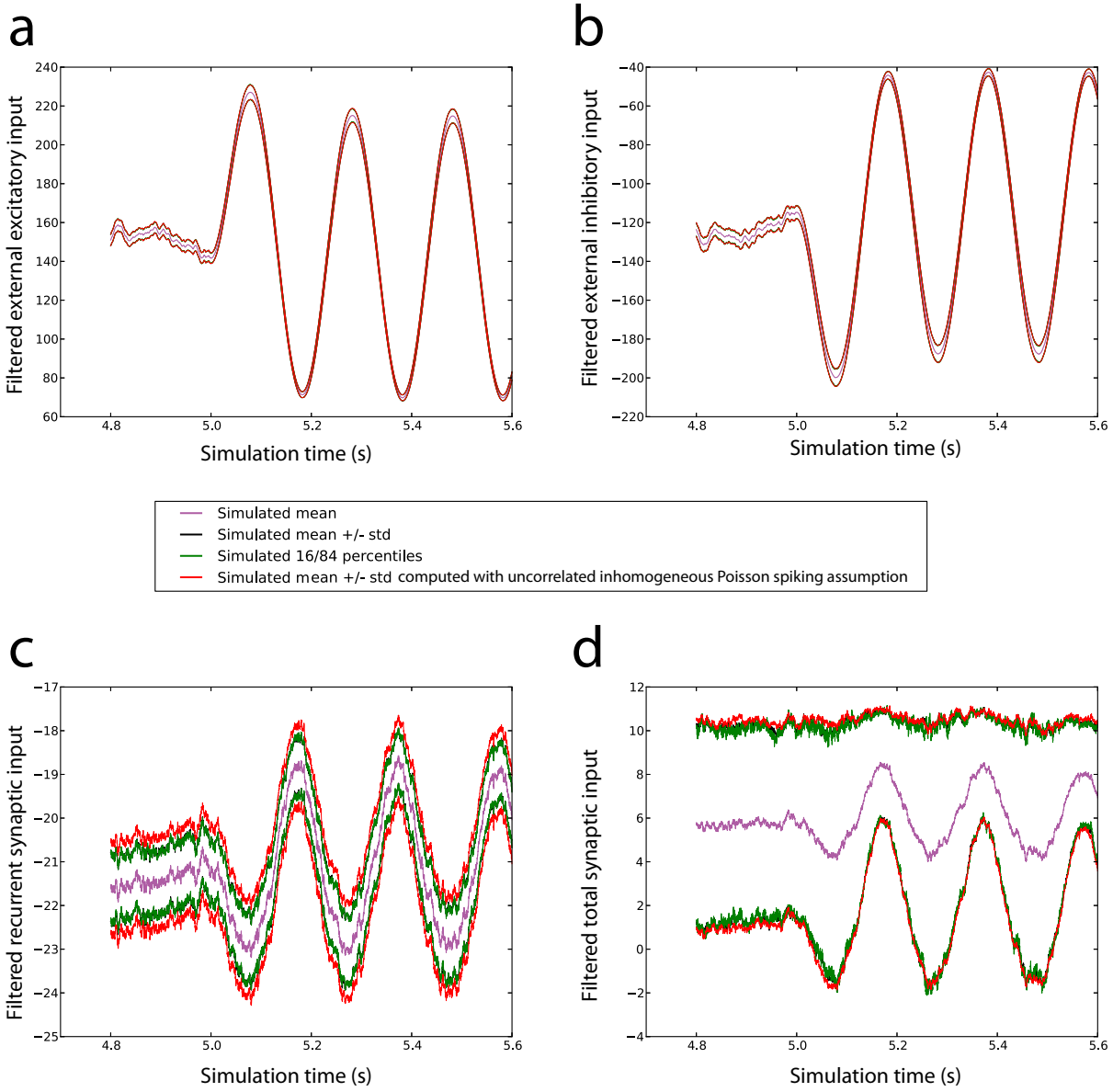


Figure 9.3: *Investigation of the shape of the distribution of filtered input in a non-stationary regime.* **(a,b,c,d)** Plot of the time-dependent characteristics of the distribution of filtered synaptic inputs. Data are shown separately for the filtered external excitatory inputs (**a**), the filtered external inhibitory input (**b**), the filtered recurrent input (**c**) and the total input $h_{i_{recc}}$ ((**d**), see the legend of Figure 9.2 and Equation 8.16 for the definitions). Note that we estimated the distribution of the variables within the network through a sample of 600 recurrent neurons (i.e., a subsample compared to the 2000 recurrent neurons). In purple, we plot the time-dependent mean filtered input as estimated from the simulation. In black, we plot two curves: one for the sum of the mean filtered input and one standard deviation of the filtered input, and another for the sum between the mean filtered input and the opposite of its standard deviation. If the distribution of filtered input were Gaussian, this +/- standard deviation confidence interval around the mean should be confounded with the 84^{th} and 16^{th} percentiles. Hence, we also plotted these 84^{th} and 16^{th} percentiles in green. Note that the black and green curves are always almost superimposed, showing that the shape of the distributions were similar to a Gaussian. Finally, we also test the time-dependent formulas for the variance that assume uncorrelated Poisson firing, and we plot in red the simulated mean filtered input +/- the square root of the variance computed through Equation 8.18. Note that for the external inputs, we use the known theoretical firing rate that we imposed as a parameter of a simulation. In contrast, for the recurrent population, we used the population-averaged firing rate recorded from the simulation (over all the 2000 recurrent neurons).

In addition, for the external filtered input, a confidence interval based on the square root of the dynamic analytical formula for the variance of the filtered input (Equation 8.18, shown in red) matched almost perfectly the values of the simulation (in black). Finally, for the recurrent filtered input, we observed again that the computation of the variance through Equation 8.18 led to an overestimation. Indeed, the confidence interval around the mean based on the square root of this semi-analytical estimate (which used the recurrent-population-averaged firing rate from the simulation, in red) was larger than the same confidence interval directly evaluated from the simulation (in black and green, see Figure 9.3 (c)). However, as described above for a steady-state simulation (Figure 9.2 (d)), this overestimation became negligible when considering the total filtered synaptic input (Figure 9.3 (d)).

More generally, we observed that, for the network parameters of our simulations, the inaccuracy due to the inhomogeneous uncorrelated Poisson firing approximation was in general small for the total synaptic input. However, this inaccuracy could become more visible in cases when the firing was strongly driven by a large increase in mean excitatory synaptic input (not shown). This was consistent with adaptation being at the origin of the overestimation of the variance of the filtered input, as spike times tend to become more strongly negatively correlated in case of large supra-threshold driving currents [Schwalger and Lindner (2013)].

Hence, we found that the uncorrelated Poisson firing approximation for evaluating the variance of the total filtered synaptic input, as well as the Gaussian approximation for the distribution of this variable within the recurrent population, gave a rather accurate description of the simulation results.

We now turn to use these approximations (as well as the other analytical considerations described in subsection 8.2.4 and subsection 8.2.5) in order to approximate self-consistently the expected firing rate within the recurrent population.

9.2 Analytical estimation of the mean firing rate within the recurrent population

We investigated the performance of our approximate analytical expressions for the expected firing rate of a neuron within the recurrent population. We used a first-moment approximation for the adaptation averaged over the intrinsic neuronal stochasticity (subsection 8.2.4). In addition, we consider a linearization of the intrinsic-stochasticity-averaged firing probability (which relates to the average firing rate over different fixed stimulations of one single neuron, subsection 8.2.5).

Using these analytical results, we reach two different expressions for the expected firing rate in the recurrent population, which correspond to two different levels of approximation:

1. An equation describing linearized, mean-input-driven, dynamics. The effects of the synaptic input variability within the population are neglected (by neglecting $\epsilon(\Delta Exc)$ in Equation 8.34, and taking the average over the populations of neurons). More specifically, we derived an expression for a linear filter Λ_{recc}, such that

$$R_{recc}(t) \approx \Lambda_{recc} * h_{recc}(t) + \left(R_{recc,\,bsln} - h_{recc,\,bsln}\left(\int \Lambda_{recc}\right)\right)$$

$$\begin{aligned} where\ h_{recc}(t) = {} & n_{ext\ exc,\,recc}\ \ F^{tot}_{ext\ exc,\,recc} * R_{ext\ exc}(t) + \\ & n_{ext\ inh,\,recc}\ \ F^{tot}_{ext\ inh,\,recc} * R_{ext\ inh}(t) + n_{recc,\,recc}\ \ F^{tot}_{recc,\,recc} * R_{recc}(t) \end{aligned} \tag{9.1}$$

In this expression, R_x is the expected population-averaged firing rate of one neuron of population x, the $F^{tot}_{x,\,y}$ are combined leak-and-synapse filters from population x to population y, and $n_{x,y}$ is the number of neurons from population x projecting to population y. In addition, $R_{recc,\,bsln}$ is the expected steady-state firing rate in a theoretical population of neurons which is matched to the simulated population in terms of mean external input, but for which there is no influence of the between-neuron variability

of the synaptic input (see section 8.2.5 for the details). Hence,

$$\begin{aligned} h_{recc,\,bsln} = n_{recc,\,recc}\, R_{recc,\,bsln} \left(\int F^{tot}_{recc,\,recc} \right) + \\ n_{ext\ exc,\,recc}\, R_{ext\ exc,\,bsln} \left(\int F^{tot}_{ext\ exc,\,recc} \right) + \\ n_{ext\ inh,\,recc}\, R_{ext\ inh,\,bsln} \left(\int F^{tot}_{ext\ inh,\,recc} \right) \end{aligned} \tag{9.2}$$

where we take for each external population the baseline rate as a time-averaged version of the external rates fed to the recurrent network: $R_{ext,\,bsln} := E_t\left[R_{ext}\left(t\right)\right]$ (see Equation 8.26). Therefore, using this approximate formula, the steady-state value of R_{recc} is $R_{recc,\,bsln}$, which is itself a non-linear function of the synaptic input (see Equation 8.27).

2. An equation approximately accounting for the non-linearity of the dynamical population response, including the effects of the variability of synaptic input within the recurrent population. Indeed, due to the non-linearity of the single neuron dynamics (see Equation 8.3), the variability of synaptic input within the neuronal population impacts the expected firing probability. In order to account for this, we preserve the exponential non-linearity for the single neuron dynamics, and we only make use of the linearization to approximate the fluctuations of the intrinsic-stochasticity-averaged adaptation (in Equation 8.38). Hence, we reach (in Equation 8.42):

$$R_{recc}\left(t\right) \propto \exp\left(E_{pop\ nrn\ recc}\left[Z_{i_{recc}}(t)\right] + \frac{var_{pop\ nrn\ recc}\left[Z_{i_{recc}}(t)\right]}{2} \right)$$

$$\begin{aligned} E_{pop\ nrn\ recc}\left[Z_{i_{recc}}(t)\right] \approx n_{ext\ exc,\,recc} \quad \left(\Phi_{ext\ exc,\,recc} * R_{ext\ exc}\right)(t) + \\ n_{ext\ inh,\,recc} \quad \left(\Phi_{ext\ inh,\,recc} * R_{ext\ inh}\right)(t) + n_{recc,\,recc} \quad \left(\Phi_{recc,\,recc} * R_{recc}\right)(t) \end{aligned}$$

$$\begin{aligned} var_{pop\ nrn\ recc}\left[Z_{i_{recc}}(t)\right] \approx n_{ext\ exc,\,recc} \quad \left(\Phi\Phi_{ext\ exc,\,recc} * R_{ext\ exc}\right)(t) + \\ n_{ext\ inh,\,recc} \quad \left(\Phi\Phi_{ext\ inh,\,recc} * R_{ext\ inh}\right)(t) + n_{recc,\,recc} \quad \left(\Phi\Phi_{recc,\,recc} * R_{recc}\right)(t) \end{aligned} \tag{9.3}$$

where the filters for computing the mean and variance of Z are such that, for any population p, $\Phi_{p,\,recc} := \left(F^{tot}_{p,\,recc} + \left(\left(\exp(\eta_{recc}) - 1\right) * \Lambda_{recc} * F^{tot}_{p,\,recc}\right)\right)$ and $\forall s,\ \Phi\Phi_{p,\,recc}(s) := \left(\Phi_{p,\,recc}(s)\right)^2$.

Hence, we can clearly see an influence of both the dynamic mean and the

dynamic variance of an effective driving input Z_{recc} (which accounts for both synaptic input and spike-driven adaptation effects) on the firing rate. As mentioned in the beginnning of this Results section, this equation includes approximations for the time-course of adaptation and for the computation of the variance of the synaptic input which tend to yield an overestimation of the firing rate in general.

We first compare our formulas with the simulation results for steady-state firing, and we then turn to non-stationary stimulation regimes.

9.2.1 Estimation of the steady-state firing rate

In Figure 9.4 (a), we compare the results of the two different levels of approximative expressions (Equation 9.1 and Equation 9.3).

Both of these expressions use an approximation on the intrinsic-stochasticity-averaged adaptation variable which is expected to underestimate its (negative) magnitude (see subsection 8.2.4). Therefore, this approximation tends to lead to an overestimation of the expected firing rate. In addition, by neglecting the effect of the variance of the synaptic input, Equation 9.1 is expected to lead to an underestimation of the drive to the network. This is due to the exponential non-linearity of the single unit's dynamics, which gives larger values for positive than for negative deviations from baseline. The two above-mentioned approximations are therefore opposite. Hence, depending on the relative magnitudes of the mean and the variance of the input, the rate predicted by Equation 9.1 may be slightly larger or slightly lower than the observed rate (Figure 9.4 (a), left vs. middle). The underestimation of the adaptation magnitude is expected to be worse in case of an increase in the negative correlations between spike times, which should be more prominent for larger mean input (see the argument in subsection 8.2.4). Accordingly, the predicted firing rates tended to be overestimated for larger mean input (Figure 9.4 (a), left vs. middle).

We will now examine the performance of the equation which approximately accounts for the effect of the synaptic input variance on the population firing rate (Equation 9.3). We recall that this equation cumulates the above-mentioned approximation for the intrinsic-stochasticity-averaged adaptation, with a linearization of the averaged adaptation time-course and an estimation of the input variance through inhomogeneous uncorrelated Poisson

firing. These three approximations are all expected to yield an overestimation of the rate. Further, this overestimation should become worse for larger firing rates. Indeed, in our simulations, Equation 9.3 consistently overestimated the firing rate, with a larger deviation for larger firing rates (Figure 9.4 (a), left vs. right). Importantly, the dependence of the firing rate on both the mean and the variance of the filtered synaptic input were still well qualitatively captured by Equation 9.3.

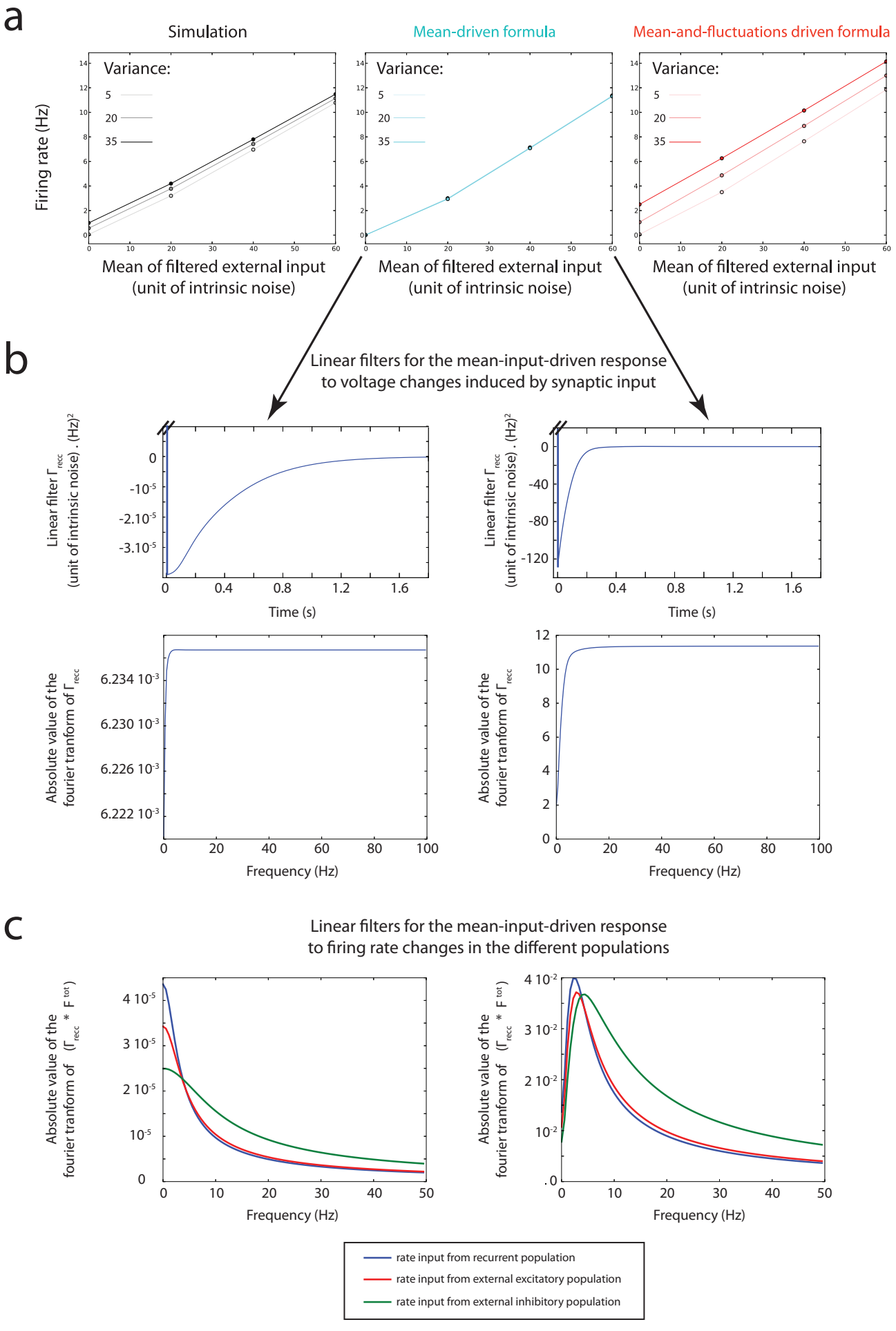

a
Simulation
Mean-driven formula
Mean-and-fluctuations driven formula
Variance:
5
20
35
Firing rate (Hz)
Mean of filtered external input
(unit of intrinsic noise)
b
Linear filters for the mean-input-driven response
to voltage changes induced by synaptic input
Linear filter Γrecc
(unit of intrinsic noise) . (Hz)²
Time (s)
Absolute value of the
fourier tranform of Γrecc
Frequency (Hz)
c
Linear filters for the mean-input-driven response
to firing rate changes in the different populations
Absolute value of the
fourier tranform of (Γrecc * F tot)
Frequency (Hz)
rate input from recurrent population
rate input from external excitatory population
rate input from external inhibitory population

Figure 9.4 *(previous page)*: *Comparison between approximate analytical expressions and simulation results for the steady-state mean firing rate within the recurrent population.* **(a)** Comparison between simulation results (different shades of grey, left), the approximate analytical formula discarding the effect of synaptic input variability on the expected firing rate within the recurrent population (different shades of blue, middle; see Equation 9.1), and the approximate analytical formula accounting for these effects (different shades of red, right; see Equation 9.3). In all cases, we compare steady-state regimes, implying that the external populations of neurons fire at constant rates. We tested different values of the mean and variance for the filtered external synaptic input I^{ext} received by a recurrent neuron i_{recc}: $I^{ext}_{i_{recc}} = \sum_{j=1}^{n_{ext\ exc,\ recc}} F^{tot}_{ext\ exc,\ recc} * S_j^{ext\ exc,\ i_{recc}} + \sum_{j=1}^{n_{ext\ inh,\ recc}} F^{tot}_{ext\ inh,\ recc} * S_j^{ext\ inh,\ i_{recc}}$. **(b)** Shape of the linear filter Λ_{recc} approximating the response of the recurrent population of neurons to changes of the mean filtered input, with a linearization around a mean filtered input of 0 (*l*eft) and around a mean filtered input of 60 (*r*ight). The filters are shown in the time domain (up), as well as in the frequency domain (bottom). **(c)** Shape of the linear filters $\Lambda_{recc} * F^{tot}$ describing the response of the recurrent population of neurons to changes of the input firing rates in the different subpopulations, in the frequency domain. These filters are valid when the effects of these input firing rate changes are largely mediated by a change in the mean filtered input received by the population (and not a change in the variability of this filtered input). The left graph is a linearization around a mean filtered input of 0, and the right graph is for a linearization around a mean filtered input of 60. We show separately:

- the filter for the response of the recurrent population to a change in the external excitatory firing rate $\Lambda_{recc} * F^{tot}_{ext\ exc,\ recc}$, in red
- the filter for the response of the recurrent population to a change in the external inhibitory firing rate $\Lambda_{recc} * F^{tot}_{ext\ inh,\ recc}$, in green
- the filter for the response of the recurrent population to a change in its own firing rate $\Lambda_{recc} * F^{tot}_{recc,\ recc}$, in blue

Finally, we examine the shape of the linear filter Λ_{recc} describing the response of the network to a delta-pulse of mean filtered firing rate in the network, which is equivalent to the approximate response of the recurrent population when all neurons are receive simultaneously the same delta pulse of h. This is also equivalent to the linearized intrinsic-stochasticity-averaged single neuron response to a deterministic filtered input (see subsection 8.2.4 and subsection 8.2.5). This filter had a delta peak at zero lag (Figure 9.4 (b), top), which reflects the fact that after the membrane and synaptic filtration, the GLM immediately responds by an increased firing rate probability to an increase in h (Equation 8.1). In addition, this initial peak was followed by a negative rebound which became more negative for the larger mean input. This reflected the effect of the adaptation variable which triggered a long-lasting decreased excitability in case of increased firing. This adaptation effect translated in the frequency domain by a high-pass filter property (Figure 9.4 (b), bottom; [Benda and Herz (2003)]). This indicates that adaptation can mitigate the effect of a slow oscillation. Indeed, if the oscillation is slow enough, then adaptation effects can develop during the rising phase of the filtered input, and then decrease when the input decreases. This can lead to a smoothing of the firing rate modulation in response to the input modulation. In addition, when considering oscillations of the firing rates (rather than oscillations of the filtered input), the effects of adaptation combine with the effects of the membrane and synaptic filters.

Typically, these filters are simple exponentials which tend to smooth fast input modulations. Hence, the total filter $\Lambda_{recc} * F^{tot}$ is a band-pass that shows a resonance (see Figure 9.4 (c)). Hence, this shows that adaptation may enhance the sensitivity of the firing rate response to the temporal structure of the input.

We now turn to the comparison between analytics and simulation during a dynamic, non-stationary regime.

9.2.2 Estimation of the firing rate in a dynamical regime

We first used a stimulation where the external excitatory population changed its rate, while the external inhibitory population fired at a constant rate. As we will see, this leads to a covariation between the mean and variance time-course for the filtered external drive. This regime is therefore favorable to the analytical expression that neglects the effect of the input variability in the population, as the time-course of the stimulation, at least, can be sensed through the mean drive.

We were also particularly interested in the dynamics induced by a change of the variability of the synaptic input within the recurrent population. Hence, we also designed a stimulation for which the mean filtered input was constant while only its variance was dynamic (as explained in subsection 8.3.3).

We will now describe to which extent our equations Equation 9.1 and Equation 9.3 could describe the mean firing rate within the recurrent population of neurons during these stimulation regimes.

Dynamical modulation of firing through correlated changes of both the mean and the variance of the synaptic input

We first examined the performance of our analytical expressions in a dynamical regime where the external inhibitory drive is constant, while the external excitatory drive varies (Figure 9.5 (a)) This led to covariations of the mean and the variance of the filtered synaptic input (Figure 9.5 (b,c)). In this regime, the expression that ignores the effect of the fluctuations (Equation 9.1) captured rather well the time-course of the averaged firing rate within the recurrent population (cyan line in Figure 9.5 (c)). This good performance probably relied in part on a compensation between the error due to the approximation of adaptation (which tends to lead to an overestimation of the

firing rate, see subsection 8.2.4) and the underestimation of the excitatory drive when neglecting the variabilty of the synaptic input in the network. In addition, as Equation 9.1 is linear, it worked better in a limited regime where the firing rate was not fluctuating a lot (left of Figure 9.5 (c)). Notably, it could not capture the clipping of the firing rate above 0, as well as the complex asymmetric shape of the rate time-course close to 0 Hz (right of Figure 9.5 (c)).

The mean-and-variance-driven, non-linear expression (Equation 9.3) captured rather well the time-course of the firing rate in all the tested stimulation regimes (red line in Figure 9.5 (c)). However, it yielded an overestimation of the firing rate that was worse for higher firing rates, for the reasons that we underlined above.

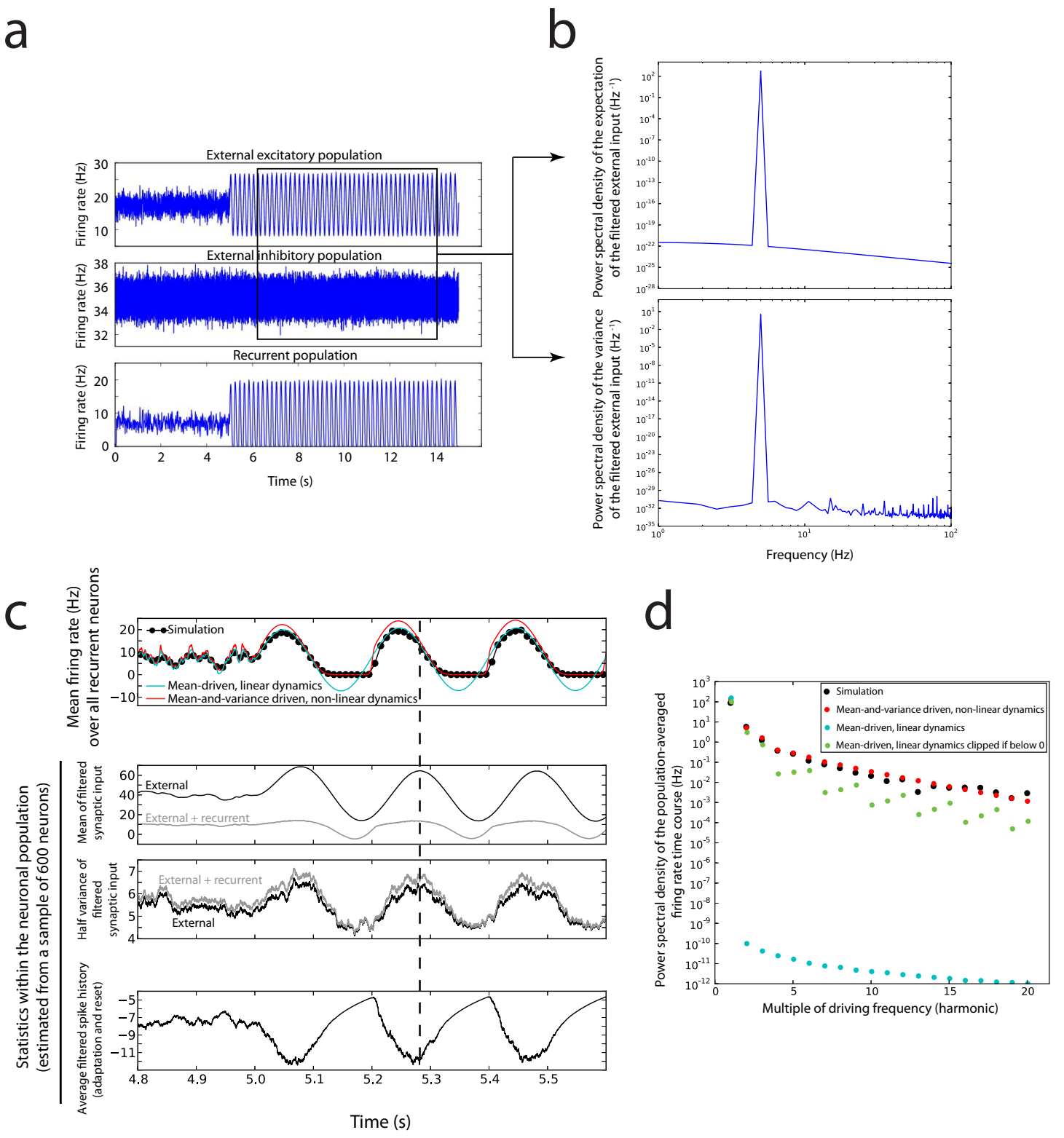

a
External excitatory population
External inhibitory population
Recurrent population
Firing rate (Hz)
Time (s)
b
Power spectral density of the expectation of the filtered external input (Hz^{-1})
Power spectral density of the variance of the filtered external input (Hz^{-1})
Frequency (Hz)
c
Mean firing rate (Hz) over all recurrent neurons
Simulation
Mean-driven, linear dynamics
Mean-and-variance driven, non-linear dynamics
Statistics within the neuronal population (estimated from a sample of 600 neurons)
Mean of filtered synaptic input
External
External + recurrent
Half variance of filtered synaptic input
Average filtered spike history (adaptation and reset)
Time (s)
d
Power spectral density of the population-averaged firing rate time course (Hz)
Simulation
Mean-and-variance driven, non-linear dynamics
Mean-driven, linear dynamics
Mean-driven, linear dynamics clipped if below 0
Multiple of driving frequency (harmonic)
e
Power spectral density of the population-averaged firing rate time course (Hz)
Simulation
Mean-and-variance driven, non-linear dynamics
Mean-driven, linear dynamics
Mean-driven, linear dynamics clipped above 0
Frequency (Hz)

Figure 9.5 ***(previous page)*:** *Comparison between approximate analytical expressions and simulation results for a dynamical regime with covariations of the mean and variance input changes.* **(a)** Mean simulated firing rate among the three populations of neurons during the whole simulation. The firing rate of the population of external excitatory neurons is time-dependent, while the external inhibitory neurons fire at a constant rate. During the first 5 seconds, the firing rate of each external excitatory neuron follows an Ornstein-Uhlenbeck process with an autocorrelation time of 5 ms, and a mean of 17.25Hz. During the following 10 seconds, we used a sine-wave of period 200ms. **(b)** Power spectral density (i.e. the Fourier transform of the autocorrelation function) for the mean and variance of the membrane-and-synapse filtered external rates during the 6.2-14.1 seconds interval. Note that we used the theoretical expected values of the rates, that we imposed, for the computation. Hence, we show the spectral content of the autocorrelation of the subthreshold membrane potential modulations induced by the external synaptic input. The top graph is for the mean $E\left[I^{ext}_{irecc}(t)\right]$, and the bottom graph is for the variance $var\left[I^{ext}_{irecc}\right]$ (see Equation 8.56). **(c)** Comparison between simulated and analytically predicted rates during a 4.6-5.8 s interval. Top: the black dotted line is the binned simulated firing rate in the whole (2000) population of neurons, the red line is the prediction by the mean-and-variance-driven, non-linear dynamics equation (Equation 9.3), and the cyan line is the prediction by the mean-driven, linear dynamics equation (Equation 9.1). Bottom: three graphs showing an estimation of the distribution of dynamical parameters in a subsample of 600 recurrent neurons (for reasons of limited computer memory). We show first the mean of the synaptic input received by each neuron (after filtering by the membrane-and-synaptic filter F^{tot}) from the external populations (black) and from all populations (including the recurrent inhibition, grey). Below, we also show half the variance of these filtered synaptic inputs within the population of recurrent neurons (in relation to the 0.5 factor in front of the variance term in Equation 9.3). We plotted a dashed line at the approximate time when the input (mean and variance) is maximal. Finally, the third graph is the mean adaptation variable in the population of recurrent neurons, clearly showing a temporal modulation on the time-scale of the firing rate modulation. **(d)** Power spectral density of the average firing rate in the recurrent population, at multiples of the driving frequency, computed within the 6.2-14.1 seconds interval. We show separately the values from the simulation (black), the values for the prediction by the mean-and-variance-driven, non-linear dynamics equation (Equation 9.3, in red), the values for the prediction by the mean-driven, linear dynamics equation (Equation 9.1, in cyan), and the values for this last prediction while negative values are clipped and set to 0 (green). **(e)** Comparison of the full power spectral densities (at all frequencies) between the data (black), and the three above-mentioned predictions. Note that the simulation values are noisy due to the finite size of the population (2000), leading to a small additional power at all frequencies.

We were interested in investigating how well Equation 9.3 actually captured the non-linear behavior of the firing rate response. Notably, there was an asymmetry of the response to a sinusoidal stimulation (Figure 9.5 (c)). This asymmetry probably resulted in part from the dynamics of the adaptation variable when the neurons were silent (Figure 9.5 (c), bottom). We also stress that another notable consequence of the adaptation dynamics was an apparent phase difference between the driving total synaptic input (both mean and variance, grey in Figure 9.5 (c), and the firing rate response. Indeed, the latter peaked before the former, and again the effect appeared to be non-linear as the firing rate peak was asymmetric.

To examine how well our expressions could capture the complex non-linear firing rate response, we examined the power spectral density (which is the fourier transform of the autocorrelation) while the sytem was responding to a

pure 5 Hz tone. Indeed, while a linear response would create a single peak at 5Hz, the response to higher harmonics is often taken as a characteristic of the non-linear response [Vasudevan et al. (2013)]. The power spectral density of the simulated average rate over the population of neurons indeed showed peaks at multiple of the driving frequency (black curve Figure 9.5 (e)). Interestingly, the mean-and-variance-driven, non-linear expression captured well the power at the higher harmonics of the response (compare red and black dots in Figure 9.5 (d); Figure 9.5 (e) left). As expected, the mean-driven, linear dynamics expression only yielded a clear peak at 5Hz (cyan items in Figure 9.5 (d-e)). Furthermore, a mere clipping of the values predicted by the linear formula above 0 (leading to a rectified linear equation, green items in Figure 9.5 (d-e)) did not fit the power at higher harmonics as well as the non-linear formula (red items in Figure 9.5 (d-e)). This probably reflects the fact that even when the negative values of the linear prediction were clipped and set to 0, the resulting rate time-course still missed the asymmetry of the firing rate time-course observed in the simulation (Figure 9.5 (c)).

Note that we also tested a non-linear dynamical formula that neglects the effects of fluctuations (a dynamic version of the equations developped in section 8.2.5). This type of equation had been briefly mentioned at the end of the discussion in Naud and Gerstner (2012a). We found that this equation could also work well for predicting the non-linear properties of the firing rate time-course in this specific type of stimulation regime, when mean and variance fluctuations are correlated (not shown). This good performance probably relied on a compensation between the underestimation of adaptation and the underestimation of the synaptic drive through ignoring the input variability. However, this compensation can only work if the mean and variance changes of the synaptic input are correlated.

Hence, we will show now that neglecting the effect of the input variance can become really problematic when the changes of input variance and the changes of input mean become decorrelated.

Dynamical modulation of firing exclusively through changes of the variance of the synaptic input within the neuronal population

In Figure 9.6, we compare our analytical expressions and the simulated average firing rate within the recurrent population, in a regime which allows to

disentangle the mean and variance effects of the synaptic input. Indeed, we tuned the inhibitory rates in order to maintain the average filtered synaptic input constant over time (Figure 9.6 (a-c), subsection 8.3.3). Note that even though such a finely-tuned stimulation is unrealistic, a balance of excitation and inhibition in biological recurrent network may be achieved by inhibitory synaptic plasticity [Vogels et al. (2011)], and could greatly moderate the changes of mean input in the network [Renart et al. (2010); Lim and Goldman (2013)]. This type of dynamical regime therefore leads to input statistics that could be close to those occurring during the stimulation regime we suggest.

By design, the mean-driven expression predicted constant firing rates over time (cyan items in Figure 9.6 (c-e)). In contrast, because of the exponential non-linearity of the single neuron input-output function Equation 8.1, when the variability of input increases within the neuronal population, the population-averaged firing rate increases (black dotted lines in Figure 9.6 (c)). In addition, even though the amplitude of the firing rate modulation that we imposed here was much more modest than in Figure 9.5, a dynamic modulation of the adaptation variable was still visible, and the population firing rate appeared to plateau before the input variance would peak. Interestingly, the mean-and-variance driven, non-linear dynamics expression appeared to capture qualitatively these features, despite the (expected) overestimation of the predicted firing rate (red curve in Figure 9.6 (c)). Finally, the power spectral density of the simulated population firing rate showed two peaks above noise level (Figure 9.6 (e)). Interestingly, the mean-and-variance driven, non-linear dynamics expression appeared to show a similar behavior (above the power of the noise that was in the simulated data, see red items in Figure 9.6 (e)).

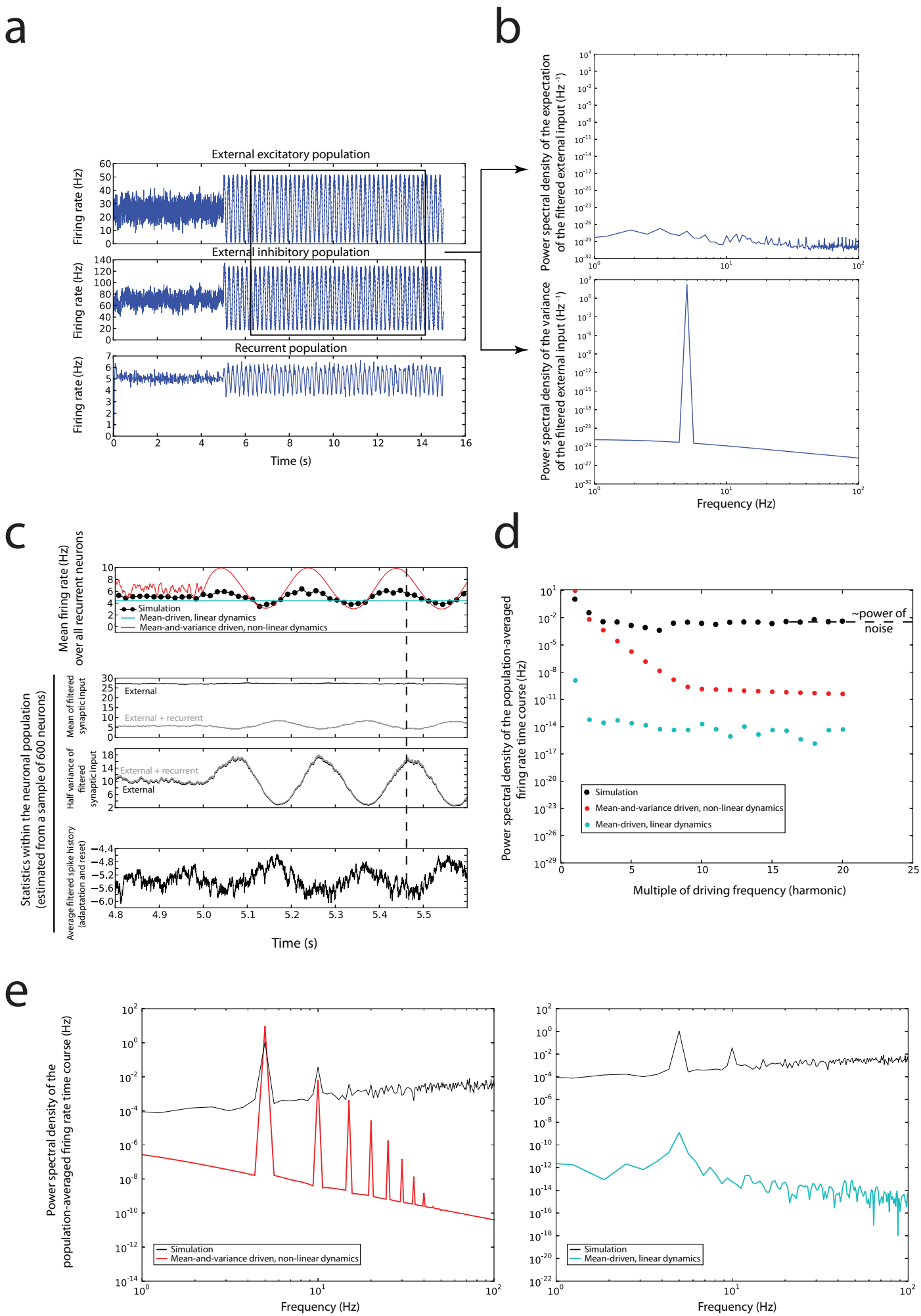

a
External excitatory population
Firing rate (Hz)
External inhibitory population
Firing rate (Hz)
Recurrent population
Firing rate (Hz)
Time (s)
b
Power spectral density of the expectation of the filtered external input (Hz⁻¹)
Power spectral density of the variance of the filtered external input (Hz⁻¹)
Frequency (Hz)
c
Mean firing rate (Hz) over all recurrent neurons
Simulation
Mean-driven, linear dynamics
Mean-and-variance driven, non-linear dynamics
Statistics within the neuronal population (estimated from a sample of 600 neurons)
Mean of filtered synaptic input
External
External + recurrent
Half variance of filtered synaptic input
External + recurrent
External
Average filtered spike history (adaptation and reset)
Time (s)
d
Power spectral density of the population-averaged firing rate time course (Hz)
~power of noise
Simulation
Mean-and-variance driven, non-linear dynamics
Mean-driven, linear dynamics
Multiple of driving frequency (harmonic)
e
Power spectral density of the population-averaged firing rate time course (Hz)
Simulation
Mean-and-variance driven, non-linear dynamics
Frequency (Hz)
Simulation
Mean-driven, linear dynamics
Frequency (Hz)

Figure 9.6 ***(previous page)*****:** *Comparison between approximate analytical expressions and simulation results for a regime where only the variability of the filtered input is dynamic.* **(a)** Mean simulated firing rate among the three populations of neurons during the whole simulation. The firing rates of excitatory and inhibitory neurons were adjusted to keep the neuron-averaged filtered external synaptic input constant, while the variability of the external input within the recurrent neuron population is dynamics subsection 8.3.3. During the first 5 seconds, the firing rate of each external excitatory neuron follows an Ornstein-Uhlenbeck process with an autocorrelation time of 5 ms, and a mean of 17.25Hz. During the following 10 seconds, we used a sine-wave of period 200ms. **(b)** Power spectral density (i.e. the Fourier transform of the autocorrelation function) for the mean and variance of the membrane-and-synapse filtered external rates during the 6.2-14.1 seconds interval. Note that we used the theoretical expected values of the rates, that we imposed, for the computation. Hence, we show the spectral content of the autocorrelation of the subthreshold membrane potential modulations induced by the external synaptic input. The top graph is for the mean $E\left[I^{ext}_{i_{recc}}(t)\right]$, and the bottom graph is for the variance $var\left[I^{ext}_{i_{recc}}\right]$ (see Equation 8.56). **(c)** Comparison between simulated and analytically predicted rates during a 4.6-5.8 s interval. Top: the black dotted line is the binned simulated firing rate in the whole (2000) population of neurons, the red line is the prediction by the mean-and-variance-driven, non-linear dynamics equation (Equation 9.3), and the cyan line is the prediction by the mean-driven, linear dynamics equation (Equation 9.1). Bottom: three graphs showing an estimation of the distribution of dynamical parameters in a subsample of 600 recurrent neurons (for reasons of limited computer memory). We show first the mean of the synaptic input received by each neuron (after filtering by the membrane-and-synaptic filter F^{tot}) from the external populations (black) and from all populations (including the recurrent inhibition, grey). Below, we also show half the variance of these filtered synaptic inputs within the population of recurrent neurons (in relation to the 0.5 factor in front of the variance term in Equation 9.3). We plotted a dashed line at the approximate time when the input (mean and variance) is maximal. Finally, the third graph is the mean adaptation variable in the population of recurrent neurons, clearly showing a temporal modulation on the time-scale of the firing rate modulation. **(d)** Power spectral density of the average firing rate in the recurrent population, at multiples of the driving frequency, computed within the 6.2-14.1 seconds interval. We show separately the values from the simulation (black), the values for the prediction by the mean-and-variance-driven, non-linear dynamics equation (Equation 9.3, in red), the values for the prediction by the mean-driven, linear dynamics equation (Equation 9.1, in cyan). **(e)** Comparison of the full power spectral densities (at all frequencies) between the data (black), and the two above-mentioned predictions. Note that the simulation values are noisy due to the finite size of the population (2000), leading to a small additional power at all frequencies.

In conclusion, while the mean-and-variance driven, non-linear dynamics expression (Equation 9.3) led to an overestimation of the firing rate, it could still capture rather well the non-linear time-course of the population-averaged firing rate in all the situations we tested.

We now turn to show how this new analytical expression can (or may be able to) clarify the neuronal mechanisms at stake during the dynamical processing implemented by the brain.

9.3 Some concrete insights reached, or probably reachable, by applying our new analytical expressions

While our new analytical expression has the disadvantage to only predict approximately the population-averaged firing rate, it also has the considerable advantage to be simple enough to provide an intuitive explanation in some concrete situations. We mention below a few of these applications, some of which having been more deeply investigated than others.

9.3.1 Log-normal distribution of the instantaneous firing rates within the population

The mean-and-variance driven, non-linear dynamics expression (Equation 9.3) is valid when three assumptions are fulfilled. First, the synaptic-input-induced membrane potential fluctuations should be Gaussian. Second, the spike time correlations –beyond the co-occurrences expected from time-dependent firing rates– should be small, or at least exert an approximately constant influence on the effective drive of the neurons (in which case our predictions would still be qualitatively correct in terms of firing rate time-course and variability of firing intensity). Third, the intrinsic-stochasticity-averaged adaptation variable should be linearizable (see section 8.2).

If these assumptions are approximately valid, our framework predicts that the instantaneous firing probability, and therefore the spike count in very small analysis windows, should follow a log-normal distribution. We could indeed observe this type of distribution in our simulations (Figure 9.7). Note that in our simulation, the log-normal distribution arose because of the interplay between the non-linearity of the single neuron input-output function, and the instantaneous variability of the synaptic input. Notably, the asymmetric distribution could occur instantaneously even though the time-averaged statistics were identical between neurons. By adding a variability in the synaptic weights received by the neurons, one could further increase the variance of the log-normal distribution (subsection 8.2.3).

Interestingly, such a log-normal distribution for the firing rate in short bins

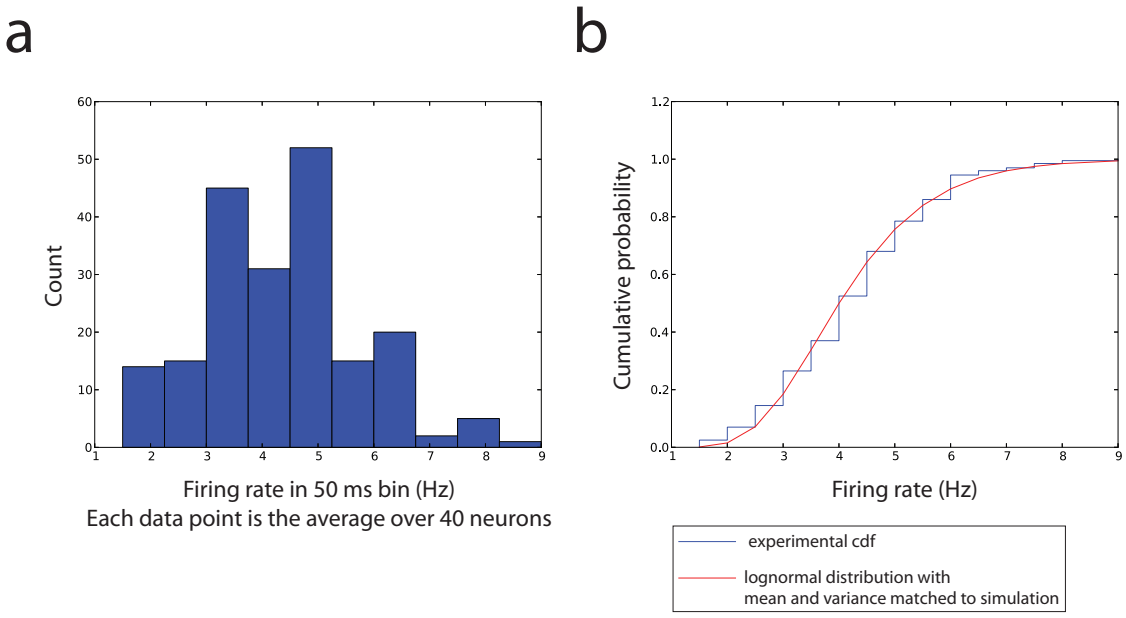


Figure 9.7: *Log-normal distribution of the instantaneous firing rate.* For this figure only, in order to increase our statistical power, we multiplied the number of recurrent neurons by 4 (hence, there were 8000 neurons), while maintaining the same number of input connections for each neurons as in previous figures. We used a steady-state stimulation during which the external filtered synaptic input (by the membrane-and-synaptic filters F^{tot}) had a mean of 20 and a variance of 35. **(a)** Histogram of the firing rate in a 50 ms bin. Each data point was the mean over 40 neurons, which is similar to the method used by [Hromádka et al. (2008)] to evaluate this instantaneous firing rate distribution. **(b)** Blue: observed cumulative distribution function (cdf) for the mean (over 40 neurons) firing rate in a 50 ms bin. In red, we plot the theoretical cumulative distribution function of a log-normal variable which mean and variance are matched to the mean and variance of the simulation.

(much shorter than the averaged inter-spike interval) was observed in the auditory cortex of awake rats [Hromádka et al. (2008)], for both spontaneous and evoked activity. This could be consistent with the idea that the approximations that we made are indeed reasonable in vivo.

Our analytical expressions may actually also yield insights into the mechanism at the origin of the changes in firing rate induced by a stimulus in vivo. Indeed, by measuring the characteristics of the distribution of instantaneous firing rate among neurons with similar response (e.g. all neurons with a similar increase of firing rate), one could deduce the mean and variance of the effective drive Z in Equation 9.3. Indeed, there is a simple relation between the mean and variance of the log-normal distribution, and the mean and variance of the underlying normal variable (that can be found back by taking the logarithm of the log-normal values). More specifically, if $X = exp(Z)$ and Z is gaussian, then $ln(E[X]) = E(Z) + \frac{var[Z]}{2}$, and $ln\left(1 + \frac{var(X)}{(E[X])^2}\right) = var[Z]$. Hence, by comparing the mean and variance of Z between spontaneous and evoked activity, one could examine whether the change in firing rate is better thought of as the consequence of a mean-driven, or of a variance-driven change in the effective driving input at the neuronal population level. This procedure may be more easily interpreted than an argument made on the variability of the inter-spike interval distribution [Compte et al. (2003); Renart et al. (2007); Mongillo et al. (2012); Deco et al. (2013)]. Indeed, this distribution is also very much influenced by the non-stationarity of the data, which may occur on time-scales that are faster than the typical inter-spike interval. Hence, a larger coefficient of variation of the interspike interval distribution may be explained either by an increase in mean-input-driven fast modulations of the firing rate, or by an increase in the instantaneous variability of the input in the population.

We note that our result can be distinguished from previous studies focusing on a log-normal distribution for the firing rates averaged over long periods of time [Roxin et al. (2011)], while we focused here on the instantaneous firing rate.

9.3.2 Speed of the population response to a change in the mean or the variance of the filtered input

Our mean-and-variance driven expression indicates that the population response to a change in the variance of the effective driving input Z occurs first

through linear filters ($\Phi\Phi$ in Equation 9.3) that are the square of the corresponding filters for the mean-driven response (Φ in Equation 9.3).

When considering a regime in which the adaptation variable is very small (e.g. for small firing rates, see Figure 9.4 (b) left), the filter Φ is merely the combined leak-and-synapse filter. This can be seen in Equation 8.39 (we remind that in this equation, the adaptation variable is captured by $\Gamma * F^{tot}$, see Equation 8.38).

Finally, in case of direct current injection at the soma of neurons during patch-clamp experiments, the combined leak-and-synapse filter can be reduced to the membrane (leaky integration) filter. This membrane filter is a simple exponential whose time-scale is equal to the membrane time-scale (see section 8.3.1). Hence, in this (overly simplified) case, Φ is a simple exponential filter with the membrane time-scale, while $\Phi\Phi$ is an exponential with half this time-scale. These filters describe the response of a population of neurons, or, alternatively, the average response of a single neuron to different stochastic current injections (as we argued in subsection 8.2.2).

As a consequence, the average time needed for a neuron to reach steady-state in response to a change in input variance would be expected to be shorter than this time in response to a deterministic step. This deterministic step in a single neuron is the equivalent to a change of population-averaged input in a neuronal population. Interestingly, several experimental studies seem to have made observations compatible with this prediction.

First, Silberberg et al. (2004) showed that an increase in the variance of an injected white input current leads to a trial-averaged response that reaches maximum faster than the response to an increase in mean current, in presence of a background synaptic input. Also, using a more realistic current with a larger autocorrelation time-scale, Tchumatchenko et al. (2011) show in their figure 3 that the plateau steady-state after a step input appears to be reached quicker for a variance change than for a mean change in the input current. This was however not quantitatively measured in this article. Also, [Tchumatchenko et al. (2011)] pointed out that when considering synaptic inputs compatible with a realistic excitatory post-synaptic potential amplitude, the response to the change in variance is much weaker in strength than the response to the mean. This is actually compatible with our observations (compare Figure 9.5 and Figure 9.6). Indeed, the filter Φ then takes values that are smaller than one, and therefore the values taken by $\Phi\Phi$ are even smaller than their square roots

(which are the values taken by Φ).

The same type of argument could be applied to better understand a recently suggested integrator network [Lim and Goldman (2013)], whose dynamics was purely variance-driven. The whole analysis of this network was made through a phenomenological firing rate model, which can be directly mapped to a spiking neuron model in case of mean synaptic input driven dynamics [Naud and Gerstner (2012a); Gerstner et al. (2014)]. Despite the fact that their phenomenological equations were valid in a very different regime than the variance-driven regime in which their spiking network was lying, the authors still found a qualitatively similar behavior in their phenomenological implementation, and in their spiking network. However, they had no quantitative prediction for the integration time-scale of the spiking network. Their phenomenological rate analysis, which could be linked to an analysis based on the filters Φ for the mean-driven dynamics, indicated that the network integration time-scale should to be proportional to the difference $(\tau_{exc} - \tau_{inh})$ between the excitatory and inhibitory synaptic time-scales. In our framework, we can consider a variance-driven network with negligible adaptation and combined leak-and-synapse filters that can be approximated by a single exponential decaying as their associated synapse (for synapses slower than the membrane time-scale, see section 8.3.1). Under these assumptions, which are likely to be valid in the spiking network of [Lim and Goldman (2013)] for low firing rates, our analysis would predict that the effective time-scale is proportional to $\frac{\tau_{exc}-\tau_{inh}}{2}$. More generally, our analytical expressions may be used to improve the understanding of the integrator network suggested in [Lim and Goldman (2013)].

9.3.3 Multiplicity of the steady-state solutions for one recurrently connected population

As mentioned in the Methods section, for a single recurrent population, the fixed-point expression for our mean-and-variance driven, non-linear dynamics framework reduces to a Lambert-W function (see Equation 8.43). Indeed, the steady-state (ss) equation reads $R_{recc,\,ss} = C \exp\left(R_{recc,\,ss}\left((\int \Phi) + \frac{\int(\Phi\Phi)}{2}\right) + I_{filt,ss}\right)$. Here, $I_{filt,ss}$ is a constant amounting to the steady-state external synaptic input filtered by the corresponding mean and variance filters (see Equation 9.3), and C and

$\left((\int \Phi) + \frac{\int(\Phi\Phi)}{2}\right)$ are other constants. The constant C must be positive. Note that this would also hold in cases when there are several recurrent populations, but only one of the populations needs to the modeled non-linearly.

The Lambert-W function (which is more formally defined as a transcendental equation) is more often defined as the values y that satisfy $y\exp(y) = x$. This is equivalent to our steady-state equation when setting $y = -R_{recc,\,ss}\left((\int \Phi) + \frac{\int(\Phi\Phi)}{2}\right)$ and $x = -C\left((\int \Phi) + \frac{\int(\Phi\Phi)}{2}\right)\exp\left(I_{filt,ss}\right)$. Interestingly, the Lambert-W function can have zero, one or two well-defined solutions depending on the value of x, which allows us to characterize the steady-state properties of the network as a function of its parameters. This is of interest as multi-stability has been proposed as a potential mechanism for cognitive processes such as memory or decision-making [Brunel and Wang (2001); Renart et al. (2007); Mongillo et al. (2012); Deco et al. (2013)]

The Lambert-W function has only one solution if $x > 0$, which translates into $\left((\int \Phi) + \frac{\int(\Phi\Phi)}{2}\right) < 0$. Hence, this is a case when the total (mean and variance) feedback is negative. Consequently, this is likely to be a stable fixed point, which would be consistent with our simulation results showing that the firing rate within the recurrent population appeared to be stabilized after some delay. In contrast, if $-\frac{1}{e} < x < 0$, there are two solutions. This corresponds to a moderately positive total feedback. Finally, if the total feedback is too positive ($x < \left(-\frac{1}{e}\right)$), there is no steady-state solution.

In addition, this analysis can actually be better visualized through the intersections of the curves $\exp(s)$ and αs. Here, s is a scaled rate: $s = \left((\int \Phi) + \frac{\int(\Phi\Phi)}{2}\right) R_{recc,\,ss}$; note that the scaling term may be negative. In addition, α is a constant: $\alpha = \frac{1}{C\left((\int \Phi) + \frac{\int(\Phi\Phi)}{2}\right)\exp\left(I_{filt,ss}\right)}$. When α is negative, s is a rate scaled by a negative value, and there is only one solution $s < 0$. When α is positive, there may be 0 or 2 solutions depending on whether α is small or large. These situations are illustrated in Figure 9.8; the number of solutions is the number of intersection points between αs and $\exp(s)$. Note that in all conditions, we do get meaningful (positive) steady-state values for the firing rate.

Finally, it is also possible to conclude that, in the case when there are two solutions, the higher-rate fixed point has an instability to slow modulations of the

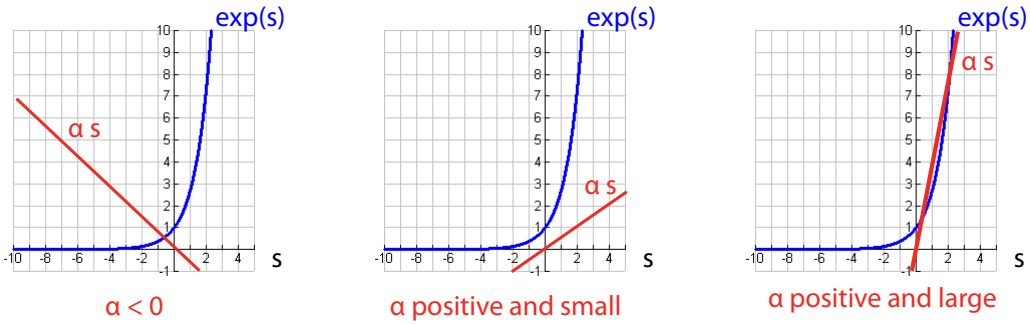


Figure 9.8: *Visualization of the steady-state solutions for one recurrently connected population.* We plot the intersections between the functions $\exp(s)$ (in blue) and $\alpha\ s$ (in red). If $\alpha < 0$, there is only one intersection (left), while if $\alpha > 0$, there may be either 0 (middle, for small α) or 2 (right, for larger α) intersections.

effective input (which includes the mean and the variance components). Indeed, in this case of slow input variations, it is possible to reason iteratively by considering that the input change leads to a firing rate change predicted by the steady-state response function, and conversely. A scaled version of this steady state rate response to an effective input s is the exponential pictured in Figure 9.8. Above the upper fixed point on the right of Figure 9.8, it is visible that if $s_{upper\ fixed\ point}$ –and therefore the effective input– increases to $s_1 > s_{upper\ fixed\ point}$, then the scaled rate increases and tends to reach a new value $R_f \approx \exp(s_1) > \alpha s_1$. After some dynamical regime, the steady-state relation indicates that s_f would tend to converge to $\frac{R_f}{\alpha}$, where s_f is the effective input received by the neurons of the network while new scaled mean firing rate of the network is R_f. Finally, $s_f \approx \frac{R_f}{\alpha} > s_1$. Hence, the rate increase leads to a large input increase, which will make the rate increase even more. In consequence, the firing rate diverges to infinity, unless another mechanism changes the steady-state single-population picture of Figure 9.8. This stabilizing mechanism could be a non-linear recurrent inhibitory current, or a non-linear adaptation threshold (which effect could be mapped onto changes of the parameters and kernels of our single-neuron model, see subsection 8.1.3).

An investigation of the stability of the other fixed points (the single fixed point at the left of Figure 9.8, or the lower fixed point at the right of Figure 9.8) would require to account for the dynamics at all time scales. This could be possible by approximating the complete filter $\left(\Phi + \frac{\Phi\Phi}{2}\right)$ through a sum of exponentials (either numerically, or, in simple cases, potentially analytically). Then, one could express the dynamics of the system as the solution of coupled non-linear differential equations (as we show in section 8.2.5).

Hence, this system may be studied with the usual stability analysis tools (linear stability, phase plane). Finally, this type of analysis could be extended to several interacting recurrent populations of neurons.

9.3.4 Modulation of the resonant frequencies for the firing rate response by adaptation

We initially started to study mean-field equations in an attempt to solve a concrete issue, as we mentioned in chapter 6. We were wondering whether one could design a temporally modulated input which would more strongly excite a population of neurons which would have undergone an episode of sustained firing in the past, compared to another population that would not have fired as much. The hypothesis was that the population of neurons which would have undergone sustained firing in the past would retain a specific adaptation state, which may interact more strongly with some types of synaptic input

First, our analysis already suggests that this type of dynamics may be possible.

Our equations indeed show that the shape of the linear filter Λ, which determines the temporal response properties of a recurrent population of neurons, can be modulated by the baseline state of the neurons (Figure 9.4 (b,c) and subsection 8.2.5). In addition, and more surprisingly, while a larger adaptation is always associated with an overall decreased excitability in a mean-driven regime, the variance-driven regime appears to be different. Indeed, the variance filter $\Phi\Phi$ is the square of the mean-driven filter Φ (see Equation 9.3). Hence, even if the linearized adaptation create a negative contribution in Φ, it will be associated with some positive terms in $\Phi\Phi$. This reflects the fact that adaptation participates to creating fluctuations. This suggests that in the variance-driven regime, the presence of adaptation currents in a population may not result in a general decreased excitability of this population.

While we did not have the time to really address this question, the framework we developed may allow us to do it in the future. Indeed, as we mentioned previously, Equation 9.3 can be reduced to differential equations of variables which may be related to the contribution of adaptation in the mean and variance filters Φ and $\Phi\Phi$. Hence, in the future, we may be able to study the influence of baseline adaptation values on the temporal properties of the neuronal population response

in the variance-driven regime.

Chapter 10

Discussion: a new tool to analyze the dynamics of recurrent adapting networks

---

We developed novel mean-field expressions for the population-averaged firing rate of recurrent neuronal networks. To the best of our knowledge, this work is the first to account for the synaptic input variability within a population of Generalized Linear Model (GLM) neurons with adaptation. By filling this gap, we connect to the existing literature investigating the different dynamical regimes that can characterize networks of other neuron models [Brunel and Hakim (1999); Lindner and Schimansky-Geier (2001); Fourcaud-Trocmé and Brunel (2005); Toyoizumi et al. (2009); Tchumatchenko and Wolf (2011); Tetzlaff et al. (2012); Helias et al. (2013); Kriener et al. (2013)]. Note that the models considered so far in the literature analyzing the variance-driven response were all non-adapting. In contrast, we use a single neuron model whose dynamics is rich enough to be fitted to recorded single neurons [Mensi et al. (2012); Pozzorini et al. (2013)] and we perform a rather detailed mathematical analysis of the population statistics in networks of interacting units, while accounting for both the response to the mean and the variance of the input in the neuronal population. Importantly, compared to most other spiking neuron models, we feel that the mathematical form of the GLM offers the advantage to allow for rather easy extensions to several important features governing the dynamics of the network. For instance, the effect of slow synaptic channels can be directly incorporated and interpreted (section 8.3.1). A linearized short-term plasticity of the synapses may also be naturally incorporated in the synaptic filters (see subsection 8.1.4 and section 8.3.1). Finally, the framework is also likely to be robust to the introduction of larger heterogeneities of the synaptic input (as argued in subsection 8.2.3).

There are still of course some limitations of the approach. First, the framework that we presented here is not trivially extensible to a variability of the adaptation parameters within a single recurrent population. This might not be too critical, however, as the data suggest that the variability of the effective adaptation properties is small within pyramidal neurons of one layer (at least among layer 5 pyramidal neurons in vitro [Pozzorini et al. (2013)]). Also, the framework cannot be easily extended to strong non-linear dendritic integration. In vitro, the non-linearity of dendrites was shown to potentially have a considerable impact on the activity, at least in layer 5 pyramidal neurons [Naud et al. (2014)]. However, in vivo, there is evidence that the constant synaptic bombardment, and the resulting high conductance state, could linearize the dendritic response [Destexhe et al. (2003)]. Such a moderate non-linearity can probably be approximated through a linear filter, at least within a some limited regime of synaptic stimulation.

Hence, we have some hope that the framework we use is relevant for describing neuronal dynamics in functional biological circuits in which neurons can be classified in groups with similar properties.

Within this framework, we contributed two novel approximate mathematical expressions.

First, we derived a simple linearized equation for mean-driven dynamics (Equation 9.1). This is, to the best of our knowledge, the first derivation of an adapting GLM's linear response function that does not require the use of statistics from a simulation, and that can therefore be computed and analyzed a priori. Indeed, in [Deger et al. (2014)], the use of a more exact and complex mathematical formalism for the adaptation yielded an expression for the linear filter which needed the simulated interspike interval distribution. Within its domain of validity, our new equation for linear mean-driven dynamics appeared to capture well the time-dependent population-averaged firing rate. More precisely, the performance is good for moderate variations of the firing rate that are induced by synaptic input dynamics which are largely determined by the changes of population-averaged filtered input (Figure 9.5). In this regime, the low-pass filtering properties, as well as the phase advance of the population-averaged response were captured (Figure 9.4 (b) and Figure 9.5 (c)). As the derivation of this expression only involves a rather simple mathematical treatment (subsection 8.2.5), we hope that in the future we will be able to better analyze and link mathematically the properties of this filter to the single

neuron parameters.

Beyond this linearized mean-driven expression, we also derive a non-linear approximate mean-and-variance driven expression for the population-averaged firing rate (Equation 9.3). This expression captured rather well the non-linear temporal response in all the various stimulation regimes we tested (see Figure 9.4, Figure 9.5 and Figure 9.6). More specifically, it could capture the asymmetry and the rectification of the rate response to sinusoidal synaptic input, as well as the apparent phase advance of the rate time-course compared to the input signal. However, this expression leads to an overestimation of the firing rate, and this overestimation becomes larger for larger firing rates. While this may be seen as a failure, we believe that the disadvantage of this inaccuracy is mitigated by the fact that we can understand where it comes from, and by the fact that we can predict when and how it will arise. Also, and perhaps more importantly, the use of this approximate mathematical expression –rather than more exact integral-equations [Naud and Gerstner (2012a); Deger et al. (2014)]– permits reaching some intuitive understanding during a few concrete situations ranging from brain recordings to complex simulations for emulating brain function. For example, the log-normal distribution of the instantaneous firing rates appears as a natural consequence of the exponential non-linearity for the single-neuron dynamics [Badel et al. (2008); Mensi et al. (2011)] and of the Gaussianity of the subthreshold membrane potential distribution [Destexhe et al. (2003)]. In addition, the simple relation (a squaring) between the linear filter for the mean and the linear filter for the variance of the effective driving input yields intuitive insights in the differences between these two dynamical regimes. More specifically, for low firing rate regimes for which the effects of the spike-history kernel can be neglected, the variance-driven stimulation appears to be governed with a time-scale that is twice faster as the mean-driven stimulation.

Furthermore, the possible reduction to transcendental equations and to differential equations potentially opens the way to using well-known tools for dynamical analysis such as visualization of the dynamics in the phase plane, and determination of the linear stability through eigenvalue decomposition. Finally, the GLM framework may also permit to interpret the neuronal dynamics in a more functional way. Indeed, thanks to the exponential non-linearity, spiking activity may be re-interpreted as a log-likelihood of, or as an information about, the stimulus dynamics that caused it [Pillow et al. (2008); Naud and Gerstner

(2012a); Park et al. (2014)].

Part IV

Conclusions

Chapter 11

Modulating the dynamics of recurrent neuronal networks by temporal signals during cognition: experimental evidence and theoretical analysis

---

In this dissertation, we exposed how we worked towards deepening the understanding of whether and how the dynamics of recurrent neuronal networks dedicated to cognitive computations could be influenced by their input's temporal structure. We argued that this question is of large interest because it relates to a basic, macroscopic property of these networks. Indeed, if these networks implement an approximate integration, they should be rather insensitive to the temporal structure of their input that is finer than their integration time-scale. In contrast, if the non-linearity of these circuits considerably shapes the result of the cognitive computation that they perform, then these networks may be considerably sensitive to their input's temporal structure.

More generally, we feel that a larger focus on the dynamics of connected neuronal populations is needed to reinforce the – still very sparse – links between theoretical and experimental work. Indeed, while different models of neuronal processing may lead to similar steady-state outcomes, their (richer) regimes of transient response are likely permit a better distinction between them. In other words, the transient response to external inputs also informs about recurrently driven dynamics, which is thought to be the basic mechanism implementing cognitive processing.

To pursue this approach, we however need in the first place to be careful to design models which are sufficiently constrained to support a non-trivial, a posteriori comparison between models and data. In addition, this requires understanding well the dynamics of realistic enough neuronal models, because the comparison has to focus on dynamical features that are truly informative about the basic mechanism characterizing the phenomenon that the model aims to explain. In contrast, we should avoid being distracted by characteristics of neuronal activity that are dependent on details of the implementation, and that are distinct from the phenomenon that one is trying to understand. Thus, these details should be ignored or at least simplified in the model (which is precisely the reason why models can be so insightful), and they should not be compared between data and model.

The approach undertaken during this doctoral study thus intended to be in line with this objective of fruitful interactions between models and data. We first evaluated the experimental evidence for a non-linear, temporally sensitive network dynamics during cognitive tasks. We then qualitatively formulated a hypothetical neuronal network mechanism that could be compatible with our observations. Finally, with the aim of progressing towards a better understanding of the network that we sketched, we worked on an approximate analytical formulation for the dynamics of recurrently connected adapting neurons. Below, we summarize the principal contributions of our work and we position them within the existing literature.

11.1 Experimental evidence for the relevance of temporal structure of cognitive signals from the dorsal Anterior Cingulate Cortex

First, we analyzed data from the dorsal Anterior Cingulate Cortex, an area which is thought to be involved in signaling the need for updating the behavioral strategy, and/or for specifying the nature of the strategy adapted to a new context [Shenhav et al. (2013)]. We focused on feedback-related discharges, which have been extensively characterized in terms of firing rate in dACC [Quilodran et al. (2008); Shenhav et al. (2013); Procyk et al. (2014)]. The area which is suspected to process these discharges, the dorsal prefrontal

cortex [Procyk and Goldman-Rakic (2006); Rothé et al. (2011); Shenhav et al. (2013)], had been shown to behave similar to a integrator in some contexts ([Kim and Shadlen (1999)], but see [Rigotti et al. (2013); Hanks et al. (2015)]). Hence, it was relevant to investigate whether dACC feedback-related discharges were likely to be decoded by an (approximate) neural integrator.

We found evidence for the functional relevance of the temporal structure of dACC spike trains at a finer resolution (τ ≈70-200ms). The optimal decoding time-scale for these temporally modulated signals was shorter than the time-scale of the firing rate response of neuronal populations (which was about 1s), and shorter than the memory time-scale required by the behavioral task (which was about 3-6s). Importantly, to the best of our knowledge, we report for the first time an analysis that goes considerably beyond a simple report of the existence of temporal structure in frontal activity. Indeed, we were careful to check the functional significance of the temporal structure. Hence, we probed whether the relative reliabilities of the spike timing and spike count signals could allow a biologically constrained and temporally sensitive decoder to extract more cognitive-control-related information than a neural integrator. We found that it was indeed the case. In addition, we reported evidence that temporal correlations and larger-than-Poisson spike count variability participated to shape the advantage of temporal structure for decoding. Furthermore, we investigated how the signals from different neurons may be combined when received by the downstream decoder. We found that a small proportion of neurons appeared to share similar temporal patterns that could complement one another during single-trial decoding. Our results also suggested that a spatial sensitivity of the decoder would allow an efficient decoding of neurons whose activity patterns are not (entirely) consistent. Finally, we extended the existing analysis methods in order to investigate the extent to which post-feedback spike timing could be predictive of the upcoming behavior of the monkey. We showed that deviations of single-neuron 1^{st} reward discharges from the prototypical, usual temporal activity pattern predicted an increased upcoming response time of the monkey. More precisely, the data suggested that for a given neuron, the deviation could occur through spike time jitters, as well as trough increased (during some trials) and decreased (during other trials) spike count. Hence, the computation of this deviation appeared to require a non-linear processing, which seems rather hard to reconcile with a decoding by a neural integrator.

Hence, altogether, our analyses bring unprecedented evidence for a temporally-

sensitive, non-linear neuronal decoder of dACC feedback-related discharges.

11.1.1 Limitations of, and questions left unanswered by, the data analysis

We did our best to make a full and rigorous use of the available data in order to test our hypotheses as well as possible, a work that required several years. However, despite our efforts, we feel that the conclusions emerging from a single data study (in general and in our particular case) cannot be regarded as definitive evidence. Of course, there is always the –human– possibility of having made a mistake during the analysis. But beyond this, there are also intrinsic limitations when analyzing a single data set. For instance, we cannot currently determine whether the slight difference between the two monkeys in the behavior-neuronal correlation merely reflects the lesser statistical robustness in one monkey, or if it may reflect some significant effect that we cannot currently interpret. Part of the difficulty therefore comes from the fact that, as for most studies using monkeys, the technical difficulties do not allow gathering data from a large number of animals. In addition, it is impossible to do an analysis without making, at times, choices whose impact on the results cannot be fully evaluated. For instance, we had the idea of testing the correlation between the deviation from prototypical spike train and behavior, but there may be another function of the neuronal activity, which we did not think of and which we did not test, that might lead to a larger correlation. If this is so, depending on the nature of this function, our interpretation of the results in terms of the characteristics of the decoder may be compromised.

Therefore, we think that the confidence level in our interpretations would be improved by a confrontation with new analyses attempting to verify the consistency of our conclusions with observations from other independent studies. For instance, if our interpretations are correct, we would expect that similar results should be observed in any context where animals must switch between different behavioral strategies. This expectation should be checked.

In addition, our study did not fully address the question of the cognitive nature of the behavioral response time modulation. Novel careful behavioral designs, with a sufficient number of trials, could try to distinguish whether the effect we measured reflected a relation of dACC discharges with the

(pre-decisional) attention magnitude, with the (post-decisional) confidence level, or with the motivation level. This would require experiments where these three factors can be decorrelated and measured (e.g. by varying independently the difficulty of the decision, and the reward received).

Moreover, our study was purely correlative. We cannot exclude that dACC activity was not causally involved in driving behavioral adaptation and response time modulations, and instead was merely reflecting the activity occurring in the causally involved area. However, designing an experiment to measure causality for such a spatiotemporal code remains an open challenge today.

Finally, our analysis could only advance a little our understanding of the precise mechanisms by which the recurrent neuronal network decoder would be impacted by dACC spike timings. A theoretical approach was therefore undertaken in an attempt to make further progress in this direction.

11.2 Theoretical analysis of the dynamic response of recurrent neuronal networks

Motivated by the question of how a temporal input (such as the one that dACC appears to send) could modulate a non-linear downstream neuronal network implementing cognitive computations, we developed new mathematical expressions to characterize the dynamical response of such networks. State-of-the-art available analysis techniques proved insufficient to address such a question. Indeed, they either used a single neuron model without adaptation which cannot reproduce well the spike timing of recorded pyramidal neurons in response to time modulated input (e.g. integrate and fire neurons, [Brunel and Hakim (1999); Kobayashi et al. (2009)]), or simplified network interactions where uncorrelated input fluctuations between neurons are ignored [Naud and Gerstner (2012a); Deger et al. (2014)]. In contrast, we used a neuronal model that can fit the dynamics of recorded neurons [Mensi et al. (2012); Pozzorini et al. (2013)], and we derived a non-linear dynamical expression which approximately accounts self-consistently for the effects of both the mean and the variability of the synaptic input. In this way, we include in the analysis the major factors governing neuronal interactions [van Vreeswijk and Sompolinsky (1996, 1998); Brunel and Hakim (1999); Renart et al. (2007); Tchumatchenko

and Wolf (2011); Mongillo et al. (2012)].

Furthermore, given the need for simple mathematical expressions in order to get intuitions in concrete applications, we undertook an effort of simplification of the existing mathematical expressions [Naud and Gerstner (2012a)]. We derived a very simple analytical linear filter for mean-input-driven adapting neuronal populations, which may be used to investigate the relation between single neuron properties and the frequency response function of the population.

We would like to reckon that, due to time limitations, the performance of our analytical expressions was not checked as extensively as we would have desired. We are aware that this will have to be done before submitting these results for publication. We are also aware that our expressions are only approximate, in particular for the amplitude of the rate response. However, we hope that, in their current state, the comparisons between simulations and our analytical expressions still show that non-trivial features of the time-course of the neuronal response are well captured in general. In addition, the simplicity of the final formulas is promising for permitting future applications to concrete neuroscience questions.

11.2.1 Future possible applications of our analytical expressions

Our mathematical expressions permit more detailed comparisons between models and data. For instance, we suggested a test to determine the contribution of the mean and the variance of the input for driving a change of activity state measured in spiking data.

In addition, the new mathematical expression suggests a possible mechanism explaining the importance of the temporal structure of dACC signals, which could reflect a tuning of this signal to the adaptation state of the decoding network. We hope to dig more into these questions in the future.

Bibliography

Abbott and van Vreeswijk C (1993). Asynchronous states in networks of pulse-coupled oscillators. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics, 48(2):1483–1490.

Alijani, A. K. and Richardson, M. J. E. (2011). Rate response of neurons subject to fast or frozen noise: from stochastic and homogeneous to deterministic and heterogeneous populations. Phys Rev E Stat Nonlin Soft Matter Phys, 84(1 Pt 1):011919.

Amit, D. J. and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. Cereb Cortex, 7(3):237–252.

Aronov, D., Reich, D. S., Mechler, F., and Victor, J. D. (2003). Neural coding of spatial phase in V1 of the macaque monkey. J. Neurophysiol., 89(6):3304–27.

Arsiero, M., Lüscher, H.-R., Lundstrom, B. N., and Giugliano, M. (2007). The impact of input fluctuations on the frequency-current relationships of layer 5 pyramidal neurons in the rat medial prefrontal cortex. J Neurosci, 27(12):3274–3284.

Avermann, M., Tomm, C., Mateo, C., Gerstner, W., and Petersen, C. C. H. (2012). Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. J Neurophysiol, 107(11):3116–3134.

Badel, L., Lefort, S., Berger, T. K., Petersen, C. C. H., Gerstner, W., and Richardson, M. J. E. (2008). Extracting non-linear integrate-and-fire models from experimental data using dynamic i-v curves. Biol Cybern, 99(4-5):361–370.

Balaguer-Ballester, E., Lapish, C. C., Seamans, J. K., and Durstewitz, D. (2011). Attracting dynamics of frontal cortex ensembles during memory-guided decision-making. PLoS Comput Biol, 7(5):e1002057.

Bekolay, T., Laubach, M., and Eliasmith, C. (2014). A spiking neural integrator model of the adaptive control of action by the medial prefrontal cortex. J Neurosci, 34(5):1892–1902.

Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P. L., Gioanni, Y.,

Battaglia, F. P., and Wiener, S. I. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal- prefrontal network upon learning. Neuron, 66(6):921–936.

Benda, J. and Herz, A. V. M. (2003). A universal model for spike-frequency adaptation. Neural Comput, 15(11):2523–2564.

Bergamaschi, L., DAgostino, G., Giordani, L., Mana, G., and Oddone, M. (2013). The detection of signals hidden in noise. Metrologia, 50(3):269276.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., and Warland, D. (1991). Reading a neural code. Science, 252(5014):1854–1857.

Blanchard, T. C. and Hayden, B. Y. (2014). Neurons in dorsal anterior cingulate cortex signal postdecisional variables in a foraging task. J Neurosci, 34(2):646–655.

Bodis-Wollner, I., Bucher, S. F., and Seelos, K. C. (1999). Cortical activation patterns during voluntary blinks and voluntary saccades. Neurology, 53(8):1800–1805.

Boucsein, C., Nawrot, M. P., Schnepel, P., and Aertsen, A. (2011). Beyond the cortical column: abundance and physiology of horizontal connections imply a strong role for inputs from the surround. Front Neurosci, 5:32.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., and Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque mt. Vis Neurosci, 13(1):87–100.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. J Comput Neurosci, 8(3):183–208.

Brunel, N., Chance, F. S., Fourcaud, N., and Abbott, L. F. (2001). Effects of synaptic noise and filtering on the frequency response of spiking neurons. Phys Rev Lett, 86(10):2186–2189.

Brunel, N. and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. Neural Comput, 11(7):1621–1671.

Brunel, N. and Latham, P. E. (2003). Firing rate of the noisy quadratic integrate-and-fire neuron. Neural Comput, 15(10):2281–2306.

Brunel, N. and Wang, X. J. (2001). Effects of neuromodulation in a cortical

network model of object working memory dominated by recurrent inhibition. J Comput Neurosci, 11(1):63–85.

Brunel, N. and Wang, X.-J. (2003). What determines the frequency of fast network oscillations with irregular neural discharges? i. synaptic dynamics and excitation-inhibition balance. J Neurophysiol, 90(1):415–430.

Buonomano, D. V. and Merzenich, M. M. (1995). Temporal information transformed into a spatial code by a neural network with realistic properties. Science, 267(5200):1028–1030.

Buschman, T. J., Denovellis, E. L., Diogo, C., Bullock, D., and Miller, E. K. (2012). Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. Neuron, 76(4):838–846.

Cain, N. and Shea-Brown, E. (2012). Computational models of decision making: integration, stability, and noise. Curr Opin Neurobiol, 22(6):1047–1053.

Carandini, M. (2012). Area v1. Scholarpedia, 7(7):12105.

Carney, L. H., Zilany, M. S. A., Huang, N. J., Abrams, K. S., and Idrobo, F. (2014). Suboptimal use of neural information in a mammalian auditory system. J Neurosci, 34(4):1306–1313.

Chase, S. M. and Young, E. D. (2007). First-spike latency information in single neurons increases when referenced to population onset. Proc Natl Acad Sci U S A, 104(12):5175–5180.

Chicharro, D., Kreuz, T., and Andrzejak, R. G. (2011). What can spike train distances tell us about the neural code? J Neurosci Methods, 199(1):146–165.

Churchland, A. K., Kiani, R., Chaudhuri, R., Wang, X.-J., Pouget, A., and Shadlen, M. N. (2011). Variance as a signature of neural computations during decision making. Neuron, 69(4):818–831.

Compte, A. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cerebral Cortex, 10(9):910923.

Compte, A., Brunel, N., Goldman-Rakic, P. S., and Wang, X. J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cereb Cortex, 10(9):910–923.

Compte, A., Constantinidis, C., Tegner, J., Raghavachari, S., Chafee, M. V.,

Goldman-Rakic, P. S., and Wang, X.-J. (2003). Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. J Neurophysiol, 90(5):3441–3454.

Deco, G., Rolls, E. T., Albantakis, L., and Romo, R. (2013). Brain mechanisms for perceptual and reward-related decision-making. Prog Neurobiol, 103:194–213.

Degenetais, E. (2002). Electrophysiological properties of pyramidal neurons in the rat prefrontal cortex: An in vivo intracellular recording study. Cerebral Cortex, 12(1):116.

Deger, M., Schwalger, T., Naud, R., and Gerstner, W. (2014). Fluctuations and information filtering in coupled populations of spiking neurons with adaptation. Phys Rev E Stat Nonlin Soft Matter Phys, 90(6):062704.

Destexhe, A., Rudolph, M., and Paré, D. (2003). The high-conductance state of neocortical neurons in vivo. Nat Rev Neurosci, 4(9):739–751.

Diba, K., Lester, H. A., and Koch, C. (2004). Intrinsic noise in cultured hippocampal neurons: experiment and modeling. J Neurosci, 24(43):9723–9733.

Dipoppa, M. and Gutkin, B. S. (2013a). Correlations in background activity control persistent state stability and allow execution of working memory tasks. Front Comput Neurosci, 7:139.

Dipoppa, M. and Gutkin, B. S. (2013b). Flexible frequency control of cortical oscillations enables computations required for working memory. Proc Natl Acad Sci U S A, 110(31):12828–12833.

Donoso, M., Collins, A. G. E., and Koechlin, E. (2014). Human cognition. foundations of human reasoning in the prefrontal cortex. Science, 344(6191):1481–1486.

Durstewitz, D., Vittoz, N. M., Floresco, S. B., and Seamans, J. K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. Neuron, 66(3):438–448.

Farkhooi, F., Froese, A., Muller, E., Menzel, R., and Nawrot, M. P. (2013). Cellular adaptation facilitates sparse and reliable coding in sensory pathways. PLoS Comput Biol, 9(10):e1003251.

Farkhooi, F., Muller, E., and Nawrot, M. P. (2011). Adaptation reduces variability

of the neuronal population code. Phys Rev E Stat Nonlin Soft Matter Phys, 83(5 Pt 1):050905.

Fenno, L., Yizhar, O., and Deisseroth, K. (2011). The development and application of optogenetics. Annu Rev Neurosci, 34:389–412.

Fourcaud-Trocmé, N. and Brunel, N. (2005). Dynamics of the instantaneous firing rate in response to changes in input statistics. J Comput Neurosci, 18(3):311–321.

Fourcaud-Trocmé, N., Hansel, D., van Vreeswijk, C., and Brunel, N. (2003). How spike generation mechanisms determine the neuronal response to fluctuating inputs. J Neurosci, 23(37):11628–11640.

Freund, T. and Kali, S. (2008). Interneurons. Scholarpedia, 3(9):4720.

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol, 61(2):331–349.

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1993). Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas". J Neurosci, 13(4):1479–1497.

Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. J Neurophysiol, 36(1):61–78.

Ganguli, S., Bisley, J. W., Roitman, J. D., Shadlen, M. N., Goldberg, M. E., and Miller, K. D. (2008). One-dimensional dynamics of attention and decision making in lip. Neuron, 58(1):15–25.

Gerstner (1995). Time structure of the activity in neural network models. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics, 51(1):738–758.

Gerstner and van Hemmen JL (1993). Coherence and incoherence in a globally coupled ensemble of pulse-emitting units. Phys Rev Lett, 71(3):312–315.

Gerstner, W. (2000). Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. Neural Comput, 12(1):43–89.

Gerstner, W. and Kistler, W. M. (2002). Spiking Neuron Models. Cambridge University Press.

Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. (2014). Neuronal Dynamics. Cambridge University Press.

Gigante, G., Mattia, M., and Del Giudice, P. (2007). Diverse population-bursting modes of adapting spiking neurons. Phys Rev Lett, 98(14):148101.

Gilmartin, M. R., Miyawaki, H., Helmstetter, F. J., and Diba, K. (2013). Prefrontal activity links nonoverlapping events in memory. J Neurosci, 33(26):10910–10914.

Gjorgjieva, J., Clopath, C., Audet, J., and Pfister, J.-P. (2011). A triplet spike-timing-dependent plasticity model generalizes the bienenstock-cooper-munro rule to higher-order spatiotemporal correlations. Proc Natl Acad Sci U S A, 108(48):19383–19388.

Goldman, M. S. (2009). Memory without feedback in a neural network. Neuron, 61(4):621–634.

Gutkin, B. S., Laing, C. R., Colby, C. L., Chow, C. C., and Ermentrout, G. B. (2001). Turning on and off with excitation: the role of spike-timing asynchrony and synchrony in sustained neural activity. J Comput Neurosci, 11(2):121–134.

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

Hanes, D. P., Thompson, K. G., and Schall, J. D. (1995). Relationship of presaccadic activity in frontal eye field and supplementary eye field to saccade initiation in macaque: Poisson spike train analysis. Exp Brain Res, 103(1):85–96.

Hanks, T. D., Ditterich, J., and Shadlen, M. N. (2006). Microstimulation of macaque area lip affects decision-making in a motion discrimination task. Nat Neurosci, 9(5):682–689.

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., and Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature, 520(7546):220–223.

Harris, K. D., Henze, D. A., Csicsvari, J., Hirase, H., and Buzsáki, G. (2000). Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. J Neurophysiol, 84(1):401–414.

Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., and Platt, M. L. (2011a). Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J Neurosci, 31(11):4178–4187.

Hayden, B. Y., Pearson, J. M., and Platt, M. L. (2011b). Neuronal basis of sequential foraging decisions in a patchy environment. Nat Neurosci, 14(7):933–939.

Helias, M., Tetzlaff, T., and Diesmann, M. (2013). Echoes in correlated neural systems. New Journal of Physics, 15(2):023002.

Hertäg, L., Durstewitz, D., and Brunel, N. (2014). Analytical approximations of the firing rate of an adaptive exponential integrate-and-fire neuron in the presence of synaptic noise. Front Comput Neurosci, 8:116.

Hoaglin, D., Mosteller, F., and Tukey, J. (1983). Understanding robust and exploratory data analysis. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley.

Holmgren, C., Harkany, T., Svennenfors, B., and Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. J Physiol, 551(Pt 1):139–153.

Hopfield, J. (2007). Hopfield network. Scholarpedia, 2(5):1977.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A, 79(8):2554–2558.

Hromádka, T., Deweese, M. R., and Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. PLoS Biol, 6(1):e16.

Huk, A. C. and Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. J Neurosci, 25(45):10420–10436.

Johansson, R. S. and Birznieks, I. (2004). First spikes in ensembles of human tactile afferents code complex spatial fingertip events. Nat Neurosci, 7(2):170–177.

Jolivet, R., Rauch, A., Lüscher, H.-R., and Gerstner, W. (2006). Predicting spike timing of neocortical pyramidal neurons by simple threshold models. J Comput Neurosci, 21(1):35–49.

Karlsson, M. P., Tervo, D. G. R., and Karpova, A. Y. (2012). Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. Science, 338(6103):135–139.

Katnani, H. A. and Gandhi, N. J. (2013). Time course of motor preparation during visual search with flexible stimulus-response association. J Neurosci, 33(24):10057–10065.

Kepecs, A., Uchida, N., Zariwala, H. A., and Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. Nature, 455(7210):227–231.

Khamassi, M., Enel, P., Dominey, P. F., and Procyk, E. (2013). Medial prefrontal cortex and the adaptive regulation of reinforcement learning parameters. Prog Brain Res, 202:441–464.

Khamassi, M., Quilodran, R., Enel, P., Dominey, P. F., and Procyk, E. (2014). Behavioral regulation and the modulation of information coding in the lateral prefrontal and cingulate cortex. Cereb Cortex.

Kim, J. N. and Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat Neurosci, 2(2):176–185.

Kobayashi, R., Tsubo, Y., and Shinomoto, S. (2009). Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold. Front Comput Neurosci, 3:9.

Koechlin, E., Ody, C., and Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. Science, 302(5648):1181–1185.

Köndgen, H., Geisler, C., Fusi, S., Wang, X.-J., Lüscher, H.-R., and Giugliano, M. (2008). The dynamical response properties of neocortical neurons to temporally modulated noisy inputs in vitro. Cereb Cortex, 18(9):2086–2097.

Kriener, B., Helias, M., Rotter, S., Diesmann, M., and Einevoll, G. T. (2013). How pattern formation in ring networks of excitatory and inhibitory spiking neurons depends on the input current regime. Front Comput Neurosci, 7:187.

Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J. Z., and Kepecs, A. (2013). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. Nature, 498(7454):363–366.

La Camera, G., Rauch, A., Lüscher, H.-R., Senn, W., and Fusi, S. (2004). Minimal

models of adapted neuronal response to in vivo-like input currents. Neural Comput, 16(10):2101–2124.

La Camera, G., Rauch, A., Thurbon, D., Lüscher, H.-R., Senn, W., and Fusi, S. (2006). Multiple time scales of temporal response in pyramidal and fast spiking cortical neurons. J Neurophysiol, 96(6):3448–3464.

Latimer, K. W., Yates, J. L., Meister, M. L. R., Huk, A. C., and Pillow, J. W. (2015). Neuronal modeling. single-trial spike trains in parietal cortex reveal discrete steps during decision-making. Science, 349(6244):184–187.

Lim, S. and Goldman, M. S. (2013). Balanced cortical microcircuitry for maintaining information in working memory. Nat Neurosci, 16(9):1306–1314.

Lindner, B. and Schimansky-Geier, L. (2001). Transmission of noise coded versus additive signals through a neuronal ensemble. Physical Review Letters, 86(14):29342937.

Litwin-Kumar, A. and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. Nat Neurosci, 15(11):1498–1505.

Liu, D., Gu, X., Zhu, J., Zhang, X., Han, Z., Yan, W., Cheng, Q., Hao, J., Fan, H., Hou, R., Chen, Z., Chen, Y., and Li, C. T. (2014). Medial prefrontal activity during delay period contributes to learning of a working memory task. Science, 346(6208):458–463.

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., and Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. Nature, 484(7394):381–385.

Llinas, R. (2008). Neuron. Scholarpedia, 3(8):1490.

Logiaco, L., Quilodran, R., Procyk, E., and Arleo, A. (2015). Spatiotemporal spike coding of behavioral adaptation in the dorsal anterior cingulate cortex. PLoS Biol, 13(8):e1002222.

London, M., Roth, A., Beeren, L., Häusser, M., and Latham, P. E. (2010). Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. Nature, 466(7302):123–127.

Luna, R., Hernández, A., Brody, C. D., and Romo, R. (2005). Neural codes

for perceptual discrimination in primary somatosensory cortex. Nat Neurosci, 8(9):1210–1219.

Lundstrom, B. N., Higgs, M. H., Spain, W. J., and Fairhall, A. L. (2008). Fractional differentiation by neocortical pyramidal neurons. Nat Neurosci, 11(11):1335–1342.

Machens, C. K., Romo, R., and Brody, C. D. (2005). Flexible control of mutual inhibition: a neural model of two-interval discrimination. Science, 307(5712):1121–1124.

Machens, C. K., Schütze, H., Franz, A., Kolesnikova, O., Stemmler, M. B., Ronacher, B., and Herz, A. V. M. (2003). Single auditory neurons rapidly discriminate conspecific communication signals. Nat Neurosci, 6(4):341–342.

Marder, E., Goeritz, M. L., and Otopalik, A. G. (2015). Robust circuit rhythms in small circuits arise from variable circuit components and mechanisms. Curr Opin Neurobiol, 31:156–163.

Martinez-Conde, S., Macknik, S. L., and Hubel, D. H. (2000). Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. Nat Neurosci, 3(3):251–258.

Martínez-García, M., Rolls, E. T., Deco, G., and Romo, R. (2011). Neural and computational mechanisms of postponed decisions. Proc Natl Acad Sci U S A, 108(28):11626–11631.

Matsumura, M., Chen, D., Sawaguchi, T., Kubota, K., and Fetz, E. E. (1996). Synaptic interactions between primate precentral cortex neurons revealed by spike-triggered averaging of intracellular membrane potentials in vivo. J Neurosci, 16(23):7757–7767.

Mease, R. A., Lee, S., Moritz, A. T., Powers, R. K., Binder, M. D., and Fairhall, A. L. (2014). Context-dependent coding in single neurons. J Comput Neurosci, 37(3):459–480.

Medalla, M. and Barbas, H. (2009). Synapses with inhibitory neurons differentiate anterior cingulate from dorsolateral prefrontal pathways associated with cognitive control. Neuron, 61(4):609–620.

Megías, M., Emri, Z., Freund, T. F., and Gulyás, A. I. (2001). Total number and distribution of inhibitory and excitatory synapses on hippocampal ca1 pyramidal cells. Neuroscience, 102(3):527–540.

Mensi, S., Naud, R., and Gerstner, W. (2011). From stochastic nonlinear integrate-and-fire to generalized linear models. In Shawe-taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K., editors, Advances in Neural Information Processing Systems 24, pages 1377–1385.

Mensi, S., Naud, R., Pozzorini, C., Avermann, M., Petersen, C. C. H., and Gerstner, W. (2012). Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms. J Neurophysiol, 107(6):1756–1775.

Michelet, T., Bioulac, B., Langbour, N., Goillandeau, M., Guehl, D., and Burbaud, P. (2015). Electrophysiological correlates of a versatile executive control system in the monkey anterior cingulate cortex. Cereb Cortex.

Moayedi, Y., Nakatani, M., and Lumpkin, E. (2015). Mammalian mechanoreception. Scholarpedia, 10(3):7265.

Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. Science, 319(5869):1543–1546.

Mongillo, G., Hansel, D., and van Vreeswijk, C. (2012). Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission. Phys Rev Lett, 108(15):158101.

Moore, J. (2007). Voltage clamp. Scholarpedia, 2(9):3060.

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. Nat Neurosci, 17(10):1410–1417.

Moreno-Bote, R. and Parga, N. (2010). Response of integrate-and-fire neurons to noisy inputs filtered by synapses with arbitrary timescales: firing rate and correlations. Neural Comput, 22(6):1528–1572.

Muller, E., Buesing, L., Schemmel, J., and Meier, K. (2007). Spike-frequency adapting neural ensembles: beyond mean adaptation and renewal theories. Neural Comput, 19(11):2958–3010.

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., and Wang, X.-J. (2014). A hierarchy of intrinsic timescales across primate cortex. Nat Neurosci, 17(12):1661–1663.

Narayanan, N. S., Cavanagh, J. F., Frank, M. J., and Laubach, M. (2013).

Common medial frontal mechanisms of adaptive control in humans and rodents. Nat Neurosci, 16(12):1888–1895.

Naud, R., Bathellier, B., and Gerstner, W. (2014). Spike-timing prediction in cortical neurons with active dendrites. Front Comput Neurosci, 8:90.

Naud, R., Gerhard, F., Mensi, S., and Gerstner, W. (2011). Improved similarity measures for small sets of spike trains. Neural Comput, 23(12):3016–3069.

Naud, R. and Gerstner, W. (2012a). Coding and decoding with adapting neurons: a population approach to the peri-stimulus time histogram. PLoS Comput Biol, 8(10):e1002711.

Naud, R. and Gerstner, W. (2012b). The performance (and limits) of simple neuron models: Generalizations of the leaky integrate-and-fire model. Computational Systems Neurobiology, page 163192.

Ohayon, S., Grimaldi, P., Schweers, N., and Tsao, D. Y. (2013). Saccade modulation by optical and electrical stimulation in the macaque frontal eye field. J Neurosci, 33(42):16684–16697.

Oram, M. W., Hatsopoulos, N. G., Richmond, B. J., and Donoghue, J. P. (2001). Excess synchrony in motor cortical neurons provides redundant direction information with that from coarse temporal measures. J Neurophysiol, 86(4):1700–1716.

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. Nat Neurosci, 17(4):594–600.

Ostojic, S. and Brunel, N. (2011). From spiking neuron models to linear-nonlinear models. PLoS Comput Biol, 7(1):e1001056.

Paiva, A. R. C., Park, I., and Principe, J. C. (2010). A comparison of binless spike train measures. Neural Computing and Applications, 19(3):405419.

Pala, A. and Petersen, C. C. H. (2015). In vivo measurement of cell-type-specific synaptic connectivity and synaptic transmission in layer 2/3 mouse barrel cortex. Neuron, 85(1):68–75.

Panzeri, S., Brunel, N., Logothetis, N. K., and Kayser, C. (2010). Sensory neural codes using multiplexed temporal scales. Trends Neurosci, 33(3):111–120.

Park, I. M., Meister, M. L. R., Huk, A. C., and Pillow, J. W. (2014). Encoding and

decoding in parietal cortex during sensorimotor decision-making. Nat Neurosci, 17(10):1395–1403.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature, 454(7207):995–999.

Pozzorini, C., Mensi, S., Hagens, O., and Gerstner, W. (2015). Enhanced sensitivity to rapid input fluctuations by nonlinear threshold dynamics. Frontiers in Neuroscience, (1).

Pozzorini, C., Naud, R., Mensi, S., and Gerstner, W. (2013). Temporal whitening by power-law adaptation in neocortical neurons. Nat Neurosci, 16(7):942–948.

Procyk, E. and Goldman-Rakic, P. S. (2006). Modulation of dorsolateral prefrontal delay activity during self-organized behavior. J Neurosci, 26(44):11313–11323.

Procyk, E., Tanaka, Y. L., and Joseph, J. P. (2000). Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. Nat Neurosci, 3(5):502–508.

Procyk, E., Wilson, C. R. E., Stoll, F. M., Faraut, M. C. M., Petrides, M., and Amiez, C. (2014). Midcingulate motor map and feedback detection: Converging data from humans and monkeys. Cereb Cortex.

Quilodran, R., Rothé, M., and Procyk, E. (2008). Behavioral shifts and action valuation in the anterior cingulate cortex. Neuron, 57(2):314–325.

Reimann, M. W., Anastassiou, C. A., Perin, R., Hill, S. L., Markram, H., and Koch, C. (2013). A biophysically detailed model of neocortical local field potentials predicts the critical role of active membrane currents. Neuron, 79(2):375–390.

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

Renart, A., Moreno-Bote, R., Wang, X.-J., and Parga, N. (2007). Mean-driven and fluctuation-driven persistent activity in recurrent networks. Neural Comput, 19(1):1–46.

Richardson, M. J. E. and Gerstner, W. (2005). Synaptic shot noise and

conductance fluctuations affect the membrane voltage with equal significance. Neural Comput, 17(4):923–947.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. Science, 306(5695):443–447.

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature, 497(7451):585–590.

Rigotti, M., Ben Dayan Rubin, D., Wang, X.-J., and Fusi, S. (2010). Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. Front Comput Neurosci, 4:24.

Rolls, E. T., Grabenhorst, F., and Deco, G. (2010). Decision-making, errors, and confidence in the brain. J Neurophysiol, 104(5):2359–2374.

Rossi, M. A., Hayrapetyan, V. Y., Maimon, B., Mak, K., Je, H. S., and Yin, H. H. (2012). Prefrontal cortical mechanisms underlying delayed alternation in mice. J Neurophysiol, 108(4):1211–1222.

Rothé, M., Quilodran, R., Sallet, J., and Procyk, E. (2011). Coordination of high gamma activity in anterior cingulate and lateral prefrontal cortical areas during adaptation. J Neurosci, 31(31):11110–11117.

Roussin, A. T., D'Agostino, A. E., Fooden, A. M., Victor, J. D., and Di Lorenzo, P. M. (2012). Taste coding in the nucleus of the solitary tract of the awake, freely licking rat. J Neurosci, 32(31):10494–10506.

Roxin, A., Brunel, N., Hansel, D., Mongillo, G., and van Vreeswijk, C. (2011). On the distribution of firing rates in networks of cortical neurons. J Neurosci, 31(45):16217–16226.

Rudolph, M. and Destexhe, A. (2003). Tuning neocortical pyramidal neurons between integrators and coincidence detectors. J Comput Neurosci, 14(3):239–51.

Rudolph, M., Pospischil, M., Timofeev, I., and Destexhe, A. (2007). Inhibition determines membrane potential dynamics and controls action potential generation in awake and sleeping cat cortex. J Neurosci, 27(20):5280–5290.

Saal, H. P., Vijayakumar, S., and Johansson, R. S. (2009). Information about

complex fingertip parameters in individual human tactile afferent neurons. J Neurosci, 29(25):8022–8031.

Sakamoto, K., Mushiake, H., Saito, N., Aihara, K., Yano, M., and Tanji, J. (2008). Discharge synchrony during the transition of behavioral goal representations encoded by discharge rates of prefrontal neurons. Cereb Cortex, 18(9):2036–2045.

Satake, T., Mitani, H., Nakagome, K., and Kaneko, K. (2008). Individual and additive effects of neuromodulators on the slow components of afterhyperpolarization currents in layer v pyramidal cells of the rat medial prefrontal cortex. Brain Research, 1229:4760.

Schall, J. D., Morel, A., King, D. J., and Bullier, J. (1995). Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. J Neurosci, 15(6):4464–4487.

Schwalger, T. and Lindner, B. (2013). Patterns of interval correlations in neural oscillators with adaptation. Front Comput Neurosci, 7:164.

Seung, H. S. (1996). How the brain keeps the eyes still. Proc Natl Acad Sci U S A, 93(23):13339–13344.

Seung, H. S., Lee, D. D., Reis, B. Y., and Tank, D. W. (2000). Stability of the memory of eye position in a recurrent network of conductance-based model neurons. Neuron, 26(1):259–271.

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

Shenhav, A., Botvinick, M. M., and Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron, 79(2):217–240.

Sheth, S. A., Mian, M. K., Patel, S. R., Asaad, W. F., Williams, Z. M., Dougherty, D. D., Bush, G., and Eskandar, E. N. (2012). Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. Nature, 488(7410):218–221.

Shmiel, T., Drori, R., Shmiel, O., Ben-Shaul, Y., Nadasdy, Z., Shemesh, M., Teicher, M., and Abeles, M. (2005). Neurons of the cerebral cortex exhibit

precise interspike timing in correspondence to behavior. Proc Natl Acad Sci U S A, 102(51):18655–18657.

Shmiel, T., Drori, R., Shmiel, O., Ben-Shaul, Y., Nadasdy, Z., Shemesh, M., Teicher, M., and Abeles, M. (2006). Temporally precise cortical firing patterns are associated with distinct action segments. J Neurophysiol, 96(5):2645–2652.

Shoelson, B. (2003). Deleteoutliers.

Silberberg, G., Bethge, M., Markram, H., Pawelzik, K., and Tsodyks, M. (2004). Dynamics of population rate codes in ensembles of neocortical neurons. J Neurophysiol, 91(2):704–709.

Sindhwani, V., Rakshit, S., Deodhare, D., Erdogmus, D., Principe, J. C., and Niyogi, P. (2004). Feature selection in MLps and SVMs based on maximum output information. 15(4):937–948.

Sompolinsky, Crisanti, and Sommers (1988). Chaos in random neural networks. Phys Rev Lett, 61(3):259–262.

Spruston, N. (2009). Pyramidal neuron. Scholarpedia, 4(5):6130.

Stevenson, I. H. and Kording, K. P. (2011). How advances in neural recording affect data analysis. Nat Neurosci, 14(2):139–142.

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

Stratanovitch, R. L. (1963). Topics in the Theory of Random Noise, volume Volume I. Gordon and Breach.

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

Szatmáry, B. and Izhikevich, E. M. (2010). Spike-timing theory of working memory. PLoS Comput Biol, 6(8).

Tchumatchenko, T., Malyshev, A., Wolf, F., and Volgushev, M. (2011). Ultrafast population encoding by cortical neurons. J Neurosci, 31(34):12171–12179.

Tchumatchenko, T. and Wolf, F. (2011). Representation of dynamical stimuli in populations of threshold neurons. PLoS Comput Biol, 7(10):e1002239.

Tetzlaff, T., Helias, M., Einevoll, G. T., and Diesmann, M. (2012). Decorrelation of neural-network activity by inhibitory feedback. PLoS Comput Biol, 8(8):e1002596.

Theodoni, P., Kovács, G., Greenlee, M. W., and Deco, G. (2011). Neuronal adaptation effects in decision making. J Neurosci, 31(1):234–246.

Thomas, E. G. F., van Hemmen, J. L., and Kistler, W. M. (2000). Calculation of volterra kernels for solutions of nonlinear differential equations. SIAM Journal on Applied Mathematics, 61(1):1–21.

Thurley, K., Senn, W., and Lüscher, H.-R. (2008). Dopamine increases the gain of the input-output response of rat prefrontal pyramidal neurons. J Neurophysiol, 99(6):2985–2997.

Totah, N. K. B., Jackson, M. E., and Moghaddam, B. (2013). Preparatory attention relies on dynamic interactions between prelimbic cortex and anterior cingulate cortex. Cereb Cortex, 23(3):729–738.

Toyoizumi, T., Rad, K. R., and Paninski, L. (2009). Mean-field approximations for coupled populations of generalized linear model spiking neurons with markov refractoriness. Neural Comput, 21(5):1203–1243.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., and Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. J Neurophysiol, 93(2):1074–1089.

Tsodyks, M. and Wu, S. (2013). Short-term synaptic plasticity. Scholarpedia, 8(10):3153.

Ullsperger, M., Danielmeier, C., and Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. Physiol Rev, 94(1):35–79.

Van Kampen, N. G. (1992). Stochastic processes in physics and chemistry. Elsevier, 2nd edition edition.

van Rossum, M. C. (2001). A novel spike distance. Neural Comput, 13(4):751–763.

Van Vreeswijk, C., Abbott, L. F., and Ermentrout, G. B. (1994). When inhibition not excitation synchronizes neural firing. J Comput Neurosci, 1(4):313–321.

van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science, 274(5293):1724–1726.

van Vreeswijk, C. and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. Neural Comput, 10(6):1321–1371.

van Wingerden, M., Vinck, M., Lankelma, J. V., and Pennartz, C. M. A. (2010). Learning-associated gamma-band phase-locking of action-outcome selective neurons in orbitofrontal cortex. J Neurosci, 30(30):10025–10038.

Vasudevan, R. K., Okatan, M. B., Rajapaksa, I., Kim, Y., Marincel, D., Trolier-McKinstry, S., Jesse, S., Valanoor, N., and Kalinin, S. V. (2013). Higher order harmonic detection for exploring nonlinear interactions with nanoscale resolution. Sci Rep, 3:2677.

Victor, J. D. (2005). Spike train metrics. Curr Opin Neurobiol, 15(5):585–592.

Victor, J. D. and Purpura, K. P. (1996). Nature and precision of temporal coding in visual cortex: a metric-space analysis. J Neurophysiol, 76(2):1310–26.

Victor, J. D. and Purpura, K. P. (1997). Metric-space analysis of spike trains: theory, algorithms and application. Network: Computation in Neural Systems, 8(2):127–164.

Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. Science, 334(6062):1569–1573.

Wainrib, G. and Touboul, J. (2013). Topological and dynamical complexity of random neural networks. Phys Rev Lett, 110(11):118101.

Wang, H., Stradtman, 3rd, G. G., Wang, X.-J., and Gao, W.-J. (2008). A specialized nmda receptor function in layer 5 recurrent microcircuitry of the adult rat prefrontal cortex. Proc Natl Acad Sci U S A, 105(43):16791–16796.

Wang, M., Yang, Y., Wang, C.-J., Gamo, N. J., Jin, L. E., Mazer, J. A., Morrison, J. H., Wang, X.-J., and Arnsten, A. F. T. (2013). Nmda receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. Neuron, 77(4):736–749.

Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron, 36(5):955–968.

Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J., and Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. Nat Neurosci, 9(4):534–542.

Wilson, H. R. and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. Biophysical Journal, 12(1):124.

Wimmer, K., Compte, A., Roxin, A., Peixoto, D., Renart, A., and de la Rocha, J. (2015). Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area mt. Nat Commun, 6:6177.

Wimmer, K., Nykamp, D. Q., Constantinidis, C., and Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. Nat Neurosci, 17(3):431–439.

Womelsdorf, T., Ardid, S., Everling, S., and Valiante, T. A. (2014). Burst firing synchronizes prefrontal and anterior cingulate cortex during attentional control. Curr Biol, 24(22):2613–2621.

Womelsdorf, T., Johnston, K., Vinck, M., and Everling, S. (2010). Theta-activity in anterior cingulate cortex predicts task rules and their adjustments following errors. Proc Natl Acad Sci U S A, 107(11):5248–5253.

Wong, K.-F. and Huk, A. C. (2008). Temporal dynamics underlying perceptual decision making: Insights from the interplay between an attractor model and parietal neurophysiology. Front Neurosci, 2(2):245–254.

Wong, K.-F. and Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. J Neurosci, 26(4):1314–1328.

Zucker, R. S. and Regehr, W. G. (2002). Short-term synaptic plasticity. Annu Rev Physiol, 64:355–405.

Appendices

List of scientific communications

Journal articles

Logiaco, L.; Quilodran, R.; Procyk, E. and Arleo, A. Spatiotemporal spike coding of behavioral adaptation in the dorsal anterior cingulate cortex. *PLOS Biology*, in press.

In preparation: A dynamic non-linear mean-field analysis of recurrent networks of adapting neurons in the asynchronous state.

Poster presentations and conference abstracts

Logiaco, L.; Deger, M.; Schwalger, T.; Arleo, A. and Gerstner W. A dynamic non-linear mean field method for networks of adapting neurons in the asynchronous state. In Beyond Mean Field workshop, Gottingen, 2015.

Logiaco, L.; Deger, M.; Schwalger, T.; Arleo, A. and Gerstner W. Towards the control of bistable attractors by temporally modulated inputs. In Bernstein Conference, Gottingen, 2014.

Logiaco, L.; Quilodran, R.; Procyk, E.; Gerstner W. and Arleo, A. Modulation of a decision-making process by spatiotemporal spike patterns decoding: evidence from spike-train metrics analysis and spiking neural network modeling. BMC Neuroscience 07/2013; 14(1) doi:10.1186/1471-2202-14-S1-P10, in Abstracts from the Twenty Second Annual Computational Neuroscience Meeting: CNS*2013.

Logiaco, L.; Quilodran, R.; Rothe, M.; Procyk, E. and Arleo, A. The spatiotemporal structure of anterior cingulate cortex activity contributes to behavioural adaptation coding. In 8th FENS Forum of Neuroscience, Barcelona, Spain, 2012.

Selected talks

04/28/2015: Seminar at Cornell University, Laboratory directed by J. Victor, NY, USA.

04/23/2015: Seminar at Columbia University, Center for Theoretical Neuroscience, NY, USA.

01/06/2015: Seminar at the European Institute for Theoretical Neuroscience

06/11/2014: Seminar at the Brain and Mind Institute research day, EPFL

07/23/2013: Seminar at the Gatsby Computational Neuroscience Unit, University College of London

Curriculum vitae of L. Logiaco

Education

2011-2015: PhD student cosupervised by A. Arleo and W. Gerstner.

Summer 2013: Advanced Course in Computational Neuroscience. Project: interaction of excitatory and inhibitory STDP in recurrrent networks.

Summer 2012: Okinawa Computational Neuroscience Course. Project: modulation of voltage-based excitatory STDP by dopamine and peri-somatic inhibition in a detailed neuron model with dendrites.

2010-2011: Second year of master degree, section cognitive sciences (summa cum laude: 'Très bien'), major in modeling - computational neurosciences ; ENS / EHESS / Universite Paris-Descartes.

2009-2010: First year of preparation to master degree of ENS/Paris 6. Classes of Neurology (From the neuron to the system ; Integrative neurosciences ; Nervous system development) and of Ecology (Evolutive genetics, Evolutive/behavioral ecology) ; options in mathematics and in physics.

2009: Graduation for the licence of life sciences of ENS/Paris 6, summa cum laude ('Très bien'). Options in mathematics and in physics.

2008-2010: supplemental classes at the Department of Cognitive Studies (DEC) of ENS: Cognitive ethology, Introduction to cognitive psychology, Introduction to neuropsychology, Introduction to computational neurosciences, Journal club in quantitative neurosciences.

2008: Admission at the competitive exams of the biology department of ENS (rank Paris: fourth), at the competitive exams 'physics–chemistry bio' (rank: first), at the competitive exams of agronomy schools (rank: third), at the competitive exams of veterinary schools (rank: first).

2006-2008: 'Classe préparatoire' (preparation to national competitive exams) section 'Biology Chemistry Physics Geology'.

2006: 'Scientific Baccalaureat' (high school diploma) option life and earth sciences (summa cum laude: 'Très bien').

Undergraduate research experience

2011: Master 2 internship in the laboratory 'neurobiology of adaptive processes', team adaptive neurocomputation, codirected by A. Arleo and E. Procyk. Thesis's title: Spike train metrics analysis of Anterior Cingulate Cortex activity in an exploration/exploitation task.

2011: Internship supervised by P. Goncalves, C. Machens and S. Deneve in the Group for Neural Theory (ENS). Top-down derived spiking neuron model for integration in oculomotor behavior; emergence of the threshold for spiking during the derivation of the cost function.

First semester 2010: Internship in the 'Auditory Neuroethology Laboratory' (University of Maryland). Acute recordings of neural activity in the superior colliculus of bats in response to 'ecological' sound sequences. Acoustic recordings of bats' vocalizations during a behavioral task (mealworm tracking). Use of these recordings as a basis for stimuli in electrophysiology experiments (extracellular multiunit activity in an awake, restrained animal).

Second semester 2009: Part time internship with Christian Machens (on the basis of a meeting every week), in the Group for Neural Theory (ENS). Model of the neural responses in macaque prefrontal cortex in a two alternative forced choice task: use of the uncertainty of the animal when the model learns the decision threshold (uncertainty measured psychophysically and electrophysiologically in rats).

Summer 2009: Internship in the Centre de Recherche en Cognition Animale, with Martin Giurfa (Toulouse). Dopaminergic blocage during appetitive conditioning in bees.