Project Number:	604102	Project Title:	Human Brain Project

Document Title:	Functional Mapping Data, Cognitive Architectures and Models for the HBP Human Brain Atlas and for First-draft HBP Brain Models: Package Two
Document Filename:	SP3 D3.7.4 FINAL.docx
Deliverable Number:	D3.7.4
Deliverable Type:	Data and models
Work Package(s):	WP3.1, WP3.2, WP3.3, WP3.4, WP3.5, WP3.6, WP3.7
Dissemination Level:	PU=Public
Planned Delivery Date:	M30/ 31 March 2016
Actual Delivery Date:	M30/31 March 2016

Authors:	Stanislas DEHAENE, CEA (P9) and SP3 task leaders
Compiling Editors:	Thiên-Ly PHAM, CEA (P9)
Contributors:	Stanislas DEHAENE, CEA (P9), Rafael MALACH, WIS (P78) Pascal FRIES, ESI (P14), Chris LEWIS, ESI (P14), Clément MOUTARD, CEA (P9), Martin GIESE, EKUT (P12), Olaf BLANKE, EPFL (P1), Nathan FAIVRE, EPFL (P1), Mel SLATER, UB (P64), Peter DE WEERD, UM (P108), Avgis HADJIPAPAS, UNIC (P84), Matias PALVA, UH (P86), Viktor JIRSA, AMU (P104), Mariano SIGMAN, CEA (P9), Rui COSTA, FCHAMP (P19), Rodrigo FREIRE OLIVEIRA, FCHAMP (P19), Tobias DONNER, UVA (P109), Andreas Karl ENGEL, UKE (P103), Talma HENDLER, TASMC (P98), Tomer GAZIT, TASMC (P98), Avi KARNI, UHAIFA (P72), Yadin DUDAI, WIS (P78), Rony PAZ, WIS (P78), Jan BORN, EKUT (P12), Lars NYBERG, UMU (P56), Johan ERIKSSON, UMU (P56), Neil BURGESS, UCL (P71), Fabian CHERSI, UCL (P71), Yves FREGNAC, CNRS (P7), Brice BATHELLIER, CNRS (P7), Thomas HANNAGAN, CEA (P9), Christophe PALLIER, CEA (P9), Florent MEYNIEL, CEA (P9), Riitta HARI, AALTO (P2), Lauri PARKKONEN, AALTO (P2), Linda HENRIKSSON, AALTO (P2)
Coordinator Review:	EPFL (P1): Jeff MULLER, Martin TELEFONT UHEI (P45): Sabine SCHNEIDER, Martina SCHMALHOLZ
Editorial Review:	EPFL (P1): Guy WILLIS, Lauren ORWIN

Abstract:	This report describes the Month 30 Deliverable for the HBP Subproject 3, Cognitive Architectures. The Deliverable, entitled “Functional mapping data, cognitive architectures and models for the HBP Human brain Atlas: package 2” aims at identifying and inventorying the data sets that SP3 is delivering at the end of the Ramp-Up Phase of the Human Brain Project. Based on this report, SP3 will make key contributions to a successful operational phase of the HBP.
Keywords:	Cognitive architectures; data; models
Available at:	www.humanbrainproject.eu/ec-deliverables

Table of Contents

Introduction

By Stanislas Dehaene and Yadin Dudai

This Deliverable identifies and inventories the theoretical syntheses, data sets, and models that Subproject (SP) 3 "Cognitive architectures" is delivering at the end of the Ramp-Up Phase of the Human Brain Project.

In this introduction, we first review what we mean by "Cognitive Architecture", how SP3 was organized, and what is being delivered here.

What is cognitive architecture?

The term "Cognitive architecture" refers to the infrastructure underlying an intelligent system: the set of internal representations, algorithms, and hardware choices that allow it to operate.

In theory, cognitive architecture might be agnostic to the hardware in which it is implemented (thus making contact with the fields of artificial intelligence and computer science). In the Human Brain Project, however, we believe that there is much to gain in analyzing the solutions that evolution implemented in the human brain. Understanding the biological hardware supporting a given cognitive architecture is a key step towards reproducing it in a machine, and so is understanding how development leads to the mature functional system. Similarly important is understanding how evolution shapes brain circuits and neural architectures, to optimize the fit between species behavior and their ecological niche. Brains have been shaped by selective pressures in which environments, capacities, biological constraints, and chance events interacted. As a result, the biological substrate, its structure, in situ activity, phylogenetic history, inter-species comparisons, and ontogenetic unfolding are all critical for understanding the brain's computational goals as well as the operation and implementation of the algorithms subserving these goals.

In a nutshell, the delineation of a "cognitive architecture" captures the brain regions and the interactions between brain regions that subserve a specific cognitive function. Defining a cognitive architecture requires the delineation of the brain areas involved, the format in which assemblies of neurons represent information, and the interconnections that allow them to exchange this information and converge onto a decision or an outcome. A reverse-engineering approach —starting from the function, analyzing it into its component parts, and then finding how these map onto brain circuits and neurons— is often useful in this endeavor.

General goals of SP3

The original aim of our subproject SP3 was to provide a survey of selected core areas in cognitive neuroscience and, for each of these areas, to critically review what is currently known of their cognitive architecture: the key principles and experimental data at the behavioural, neural, and network levels that need to be considered and incorporated into any realistic theoretical model of the brain.

The Ramp-Up Phase was focused on a specific subset of well-defined, challenging cognitive domains, already partially studied by cognitive neuroscience. For each such domain, the scientists in SP3 reviewed the existing literature and also generated data from innovative strategic experimental protocols aimed at dissecting the associated patterns of brain activation and response dynamics. The generated data aimed to provide fundamental constraints on any attempt at modelling the corresponding function. By providing such top-down constraints, arising from high-level knowledge of computational constraints, behaviour and brain circuits, cognitive neuroscientists in the HBP aimed to either create, or at least constrain, theoretical models and computer simulations that capture and reproduce the main facts about a cognitive architecture.

The selected cognitive domains were:

- Low-level perception and multimodal integration (WP3.5)
- The cycle leading from perception to action (WP3.1)
- Motivation, Decision and Reward (WP3.2)
- Learning and Memory (WP3.3)
- Core knowedge for Space, Time and Number (WP3.4)
- Capabilities characteristic of the Human Brain (WP3.6)

Each of these domains comprises, of course, a vast set of questions. In the Ramp-Up Phase, given the limited funding available, the teams focused on a narrower set of issues (further detailed in their specific subsection, as described below). For instance, within "core knowledge", it was decided to focus entirely on the representation of space (place- and grid-cell systems allowing for spatial navigation, group led by Neil Burgess), thus leaving the representations of time and number for future work beyond the Ramp-Up Phase.

Unfortunately, internal conflicts in HBP disrupted this plan. As a consequence of the initial dismissal of SP3, followed by reintegration of "systems and cognitive neuroscience" through an open call, virtually none of the scientists involved in the Ramp-Up Phase remain present in the next phase of HBP (most of them did not reapply voluntarily). Thus, the 10-year plan that was initially proposed will not be fully achieved. Still, the Ramp-Up Phase led to a very significant productivity, with more than 30 published peer-reviewed scientific papers, 19 new data sets and 5 novel models developed.

Description of the Deliverables

Reviews. For each cognitive function or sub-function under study, the researchers in SP3 developed detailed reviews of what they consider the essential facts about circuitry, physiology and function that any neuronal model of the corresponding cognitive processes should reproduce. These reviews have all been included in the present document.

To improve the impact on the neuroscience community, the vast majority of these reviews were published in a special issue of the journal *Neuron* (published on October 7, 2015; co-edited by Stanislas Dehaene (CEA) and Yadin Dudai (WIS), with help from Katja Brose and Christina Konen at *Neuron*). This issue contained a total of 15 papers on "Cognitive Architectures", the majority of which were authored by members of SP3. The special issue was distributed broadly at the Society for Neuroscience.

We believe that this publication is an important achievement of the present SP3, as it indicates that our central Deliverables have been peer-reviewed and passed the publication stage. Due to space or timing issues, a few reviews included in the present document have not been published yet, but their authors intend to submit them for publication too.

Data and models.

As detailed further below, SP3 is delivering the following set of data and models:

DATA

1.5 Dynamics of the internal model of objects and faces (human behaviour and MEG)

1.5 Localization and dynamics of spontaneous activity in visual areas (human fMRI)

1.3 Dynamics of attention (non-human primate local-field potentials)

1.7 Cortical representation of the body (human behaviour, ERPs and fMRI)

1.4 Map of human inter-areal connectivity and phase lags (human SEEG)

2.1 Human networks involved in computing confidence (human behaviour and fMRI)

2.1 Mouse computation of confidence in action (mice behaviour)

2.2 Human networks involved in motivation and effort (human behaviour and fMRI)

2.3 Brainstem modulation of decision processes (human behaviour and fMRI)

2.4 Human networks for motivation, decision and valuation (intracranial recordings)

3.1 Brain signatures of procedural memory encoding and consolidation (human behaviour and fMRI)

3.1 Human networks for episodic memory encoding and consolidation (human behaviour and fMRI)

3.2 Human network for conscious and unconscious working memory (human fMRI)

1.1 Unisensory and Multisensory integration in primary sensory cortices in rodents and higher mammals

1.1 Database of neuronal recordings in primary visual cortex (cat intracellular data)

1.1 Neural responses to unimodal and multi-modal stimuli (mice two-photon data)

5.2 Human networks encoding syntactic structures (human fMRI)

5.2 Cortical encoding of probabilistic sequences (human behaviour and MEG)

5.2 Brain networks encoding geometrical sequences (human and monkey fMRI)

5.3 Human networks for social cognition (human fMRI)

MODELS

1.2 Model of gamma oscillations in visual cortex

1.6 Model of visual action recognition (+ stimuli + human behavioural data)

3.1 Neural mass model of the sleeping brain

4.1 Model of spatial navigation and spatial memory

5.1 Model of the emergence of human areas responsive to letter and number symbols

The corresponding experiments and simulations are described in detail in the following sections of this document.

The SP3 coordinator (CEA, directed by S. Dehaene) ensured that all researchers filled in the Dataset Information Cards, which provide detailed information about the data, the **location of the data storage** and the **provenance of the data**. All data is documented and

is available for download, either on the local institution's server, or on the FTP site provided by HBP.

Self-assessment of the outcome of the Ramp-Up Phase

Publications resulting from this work. The research performed by SP3 has resulted in more than 30 publications, the vast majority of which appeared in top-ranking journals (*Neuron, Trends in Cognitive Science, Current Biology*, *The Journal of Neuroscience*, etc.). In this document, a list of publications is included in each section. Several additional publications are in preparation.

Indication of who has used this work so far, and for what. Most theoretical syntheses produced by SP3 researchers have been published in the form of papers available to the whole scientific community and, in many cases, already cited.

The primary use of the experimental data and models has been the normal scientific process of generating publications and presentation and discussion in seminars and international meetings.

Furthermore, throughout the Ramp-Up Phase, several **collaborations and interactions** with other subprojects and within SP3 took place. Here are a few examples:

- *WP3.1 Perception-Action*

Pascal Fries' group (ESI) collaborated with Gustavo Deco's group (SP4) on modelling ECoG. Martin Giese's group (EKUT) interacted with SPs 4 and 5 on spiking neuron models (Grün / Diesmann). Peter De Weerd (UM) and Avgis Hadjipapas (UNIC) discussed collaborations with Gaute Einewoll (SP4) and Markus Diesmann (SP6) on biophysical/hybrid models of cortical columns.

- *WP3.2 Motivation, Decision and Reward*

Mariano Sigman (CEA/University of Buenos Aires), Florent Meyniel (CEA) and Tobias Donner (UVA) initiated a collaboration to investigate confidence in a perceptual decision.

- *WP3.3 Learning and Memory*

Lars Nyberg's group (UMU) collaborated with Anders Lansner (SP9) and Ed Vogel (external) when producing the cognitive architecture description.

- *WP3.4 Space, time and numbers*

Neil Burgess's group discussed with SP4 at the EITN on simulations of navigation, and with Gustavo Deco on MEG data relating to large-scale network models of brain function. Collaborations are planned with the new SP3 project EPISENSE (PI Cyriel Pennartz).

- *WP3.5 From sensory processing to multimodal perception*

Yves Frégnac's group interacted with SP4 and the EITN Institute.

- *WP3.6 Capabilities characteristic of the human brain*

Thomas Hannagan (CEA) had positive interaction with Marc de Kamps (SP4) leading to a possible collaboration. Lauri Parkkonen (AALTO) has participated in the meetings of the "Magnetrodes" EU FP7 project, which aims at measuring neuromagnetic fields at single-neuron scales.

Making the data available via HBP platforms.

A strong effort was made to meet the HBP goal of integrating key data and making it available to others . All data was documented in detail and either loaded onto the website as instructed by our SP5 contact (Martin Telefont), or in a few cases made available on an institutional server. Detailed single-subject data, ready for further analysis, has been provided (e.g. reference data on human brain responses to 35 types of syntactic constructions, Pallier et al [see below]). Furthermore, most models developed in SP3 have been made available in the format appropriate for HBP (e.g. Martin Giese re-developed his simulations in the simulator NEST, one of the official HBP simulation tools).

However, at the time of this writing, the SP3 data could not be made available publicly through the HBP platforms. This is because the platforms are officially opening at the end of March 2016. Throughout the Ramp-Up Phase, there was no platform on which we could deposit the data. Furthermore, the pilot platforms are not yet as user-friendly as initially envisaged at the beginning of the project. In particular, creating *de novo* adequate metadata structures for system and cognitive neuroscience compatible with the future HBP database structures would have required allocation of resources which were not available or planned during the Ramp-Up Phase.

Since current SP3 members are leaving the HBP at the end of the Ramp-Up Phase, the new HBP teams will be in charge of future data integration and release, with the goal of fostering additional scientific collaborations.

Self-analysis of the value and completeness of the data.

As the present report as well as the quality of the publications demonstrates, SP3 has delivered highly valuable, detailed and useful syntheses, new reference data, and new models of specific cognitive architectures. Still, a number of caveats should be emphasized.

Data coverage and "completeness". The data that were acquired obviously correspond to a small subpart of the data needed to characterize a given cognitive function. In the limited time available (2.5 years), the SP3 leaders endeavored to design cognitive paradigms that would provide new reference data that they considered essential for any scientific description of the corresponding function. Given the infinite space of cognitive stimulation paradigms, we do not fully understand the question, raised by the referees, of the "completeness" of the data: unlike, say, cell counts or receptor concentration data, there is no clear point at which such data would be "complete". However, the data are clear "complete" in the sense that all planned data acquisition have been completed and, in most cases, the data have been fully analyzed and are either published or in preparation for publication.

Limited funding. The SP3 "Cognitive architecture" received a total of 495 person x months (PM) for 2.5 years and for 19 tasks, i.e. less than one full-time person per task on average. Money was not equally distributed, with some partners receiving 6 PM while others (particularly those joining through open-calls) receiving much more. For most teams, this was barely enough to cover the salary of a post-doc, plus scanning costs and the travel costs needed to participate in the many meetings imposed by HBP administration. The SP3 achievements listed in the present document should be considered in proportion to this funding level, in particular in the case of system neuroscience where animal costs and the

development of specialized technological equipment have to be supported. As an example, the acquisition of new multiscale data on the early visual system of higher mammals (cats and monkeys) and primary cortices of rodents (e.g. task 3.1.1 and WP3.5) were funded almost entirely by sources other than HBP, although HBP support was mentioned in the acknowledgements of the published or in press papers.

Over-ambitious description of work (DoW). For some groups, the description of work that would be performed in the Ramp-Up Phase was over-ambitious. This is the case in particular for the work on action perception, where the intended program consisted in integrating reference data into a working simulation capable of reproducing the main phenomena in the field, not in gathering a complete set of fMRI, M/EEG, intracranial and electrophysiological responses. The fact that HBP hired science writers to rewrite the DoW, without necessarily consulting the scientists on the proposed changes, is in great part responsible for this misunderstanding. The DoW was written from documents that corresponded to the ambitious 10-year HBP program rather than to what could be reasonably accomplished in the 2.5 year ramp-up, especially given the available funds. Nevertheless, as described below, each group completed its task and fulfilled its goal of providing, for a given cognitive functions, the key facts that any future brain-modelling project should reproduce.

Impossibility of performing the non-human primate research part of the program. At the start of the project, primate researchers in SP3 were dismayed to discover that the ethical form of the HBP agreement had been submitted with a clear mention that no new data would be acquired in non-human primates – although such data were indispensable to the proposed research program (e.g. WP3.6 Characteristics of the human brain). This disagreement led to the immediate resignation of one scientist (Andreas Nieder) and the reorientation of the corresponding program, with only 1.5 years left, to a more limited reviewing and modeling role (work by Thomas Hannagan, described below). Other groups (e.g. Pascal Fries, Liping Wang and Stanislas Dehaene) collected data using other sources of funding, and used the limited HBP funding to organize, analyze and especially theorize the data. In addition, the group of Peter De Weerd limited itself for the NHP data analysis part to NHP data collected prior to obtaining HBP funding.

Cognitive architectures for Perception and Action

WP 3.1 coordinated by Olaf Blanke

and WP 3.5 coordinated by Yves Frégnac

This Work Package WP3.1 is led by Olaf Blanke (EPFL) and involves five tasks on non-consious and conscious visual recognition (T3.1.1), circuits linking perception to action (T3.1.2), body perception and sense of self (T3.1.3), multiscale data analysis and transfer modeling (T3.1.4) and physiologically constrained brain network models (T3.1.5).

Work in T3.1.1 by WIS investigates human intra-cranial recordings (ECoG) in a visual categorization task and revealed transient short latency visual and motor responses in frontal and parietal cortex, suggesting that working memory encoding is achieved by transiently activating slow synaptic processes. Work by ESI investigates the relationship between spontaneous vs. stimulus evoked inter-areal interactions with ECoG recordings from non-human primates. This work showed that both intrinsic and evoked visual responses are characterized by a high degree of spatial and spectral specificity. Relying on MEG/EEG recording, work by CEA documented the fundamental laws of invariant object recognition as well as the decoding of rotating mental images. Work by EKUT in T3.1.2 investigates neural models for the recognition of actions, and specifically goal-directed actions. A neurodynamic model for the perception of body motion (observed from multiple views) was developed, exploring probabilistic models of action semantics using Markov Logic. Relying on virtual reality and on MR-compatible robotic technology, work by EPFL in T3.1.3 investigates multisensory mechanisms of illusory own body perceptions concerning the hand and the full-body in conjunction with behavioral and neuroimaging (fMRI) analysis. Along the same lines, work by UB in T3.1.3 investigates multisensory mechanisms of illusory own body perceptions concerning the hand and the full-body. For this combined virtual reality, behavioral, physiological and high-density EEG data were recorded in humans to establish electrophysiological measures for hand and full-body ownership.

In addition to these three main tasks, two new projects focusing on methodological development started in April 2014. Work by UM and UNIC in T3.1.4 developed new tools for multi-scale data analysis and multi-scale transfer modeling, linking LFP, ECoG, and MEG data. Work by UH and AMU developed new tools to describe the spatial and temporal structure of the brain with MEG and SEEG.

Altogether, this Work Package aims at providing a spatial and temporal description of neuronal circuits implicated in specific well-characterized cognitive task, and determining a list of specific constraints on human brain modeling for these selected cognitive functions.

The workpackage WP3.5 is led by Yves Frégnac (CNRS).

The dominant feedforward view of visual processing (Hubel and Wiesel, 1962; 1968) is based on the repetition, at each stage of integration, of canonical, but highly specific, rules of anatomical convergence from which derives the function. Although this simplifying view of sensory processing has led to major advances, it fails to account for the functional complexity expected from the recurrent structural connectivity of cortical subcircuits on the one hand, and the non-linear nature of the dynamic interactions between excitation and inhibition during sensory processing. Furthermore, most of understanding has been established using highly standardized and parametrized sensory contexts (spots, bars and gratings), which have little to do with the rich spatio-temporal statistics experienced during the natural scene viewing conditions of our everyday life. In addition, the main conceptual limit to our present knowledge of early visual processing is that most modeling efforts have been targeted at explaining sensory discharges only at the spike level in a purely phenomenological perspective (see Carandini et al, 2005 for a review) rather than aiming at elucidating causal conductance-based mechanisms regulating the temporal selectivity of the spiking opportunity window.

Using advanced intracellular and 2-photon imaging techniques, we have studied in the cat and mouse visual cortex generic principles of sensory processing and identified plausible neuronal correlates of Gestalt Laws and Multimodal Perception.

The first task (3.5.1) focuses on the visual processing in the early visual system of the anesthetized cat. The choice of this preparation is dictated by the need to control precisely in space and time the reproducibility of the visual input, and ensure mechanical stability necessary for long duration recordings. In the first task of WP3.5.1, we have documented to an unprecedented level multiscale dynamic states evoked by standardized visual input, ranging from Dirac, dense noise and Fourier inputs to more realistic natural-like statistics. For the exact same seed of stimulus, a comparative description of sensory responses is given, ranging from conductance and intracellular evoked waveforms, to multi-unit and local field potential dynamics. The main result is that the sensory code and its temporal precision are stimulus-dependent and optimized by natural input statistics. In contrast to Fourier inputs, broad-band spectrum stimuli elicit synchronized input from the “silent surround” of V1 receptive fields with a significant alpha-band contribution

In the second task of WP3.5.1, we show how reverse engineering approaches can be used to establish causal links between the functional dynamics of synaptic echoes in primary visual cortex and perceptual biases in low-level non-attentive perception. In particular, we demonstrate the existence of combined representation in V1 of synaptic integrative mechanisms facilitating the binding between orientation co-linearity and global motion flow on the one hand, and common fate detection on the other hand. Some of these processes seem particularly adapted to integrate the visual flow during saccadic eye-movements, when the feedforward flow of retinal input is in phase with intracortical horizontal propagation. We propose that they participate to a dynamic reconfiguration of the association field of visual cortical neurons during oculomotor scanning of natural scenes.

The second task (WP3.5.2) shifts from the cat experimental model to the mouse model, judged to be optimal to study synaptic and functional interactions across primary sensory areas and reveal the neural basis of multimodal cortical intégration. Using 2-photon imaging in the behaving mouse, we demonstrate that functional excitatory interactions lead to a sparse representation of sound features in the visual encoding space. This representation adds up mostly linearly to the visual representation but some even sparser nonlinear cells might encode specific combination of auditory and visual features.

1.1 Early sensory processing: unimodal and multimodal responses

WP3.5 – Yves Frégnac (CNRS-UNIC Unité de Neurosciences Information et Complexité)

Review on the cognitive architecture for early sensory processing

Yves Frégnac and Brice Bathellier "Cortical Correlates of Low-Level Perception: From Neural Circuits to Percepts", **Neuron**, Volume 88, Issue 1, p110–126, 7 October 2015

Extended summary:

Low-level perception results from neural-based computations on sensory information, and serves to build unconscious or self-generated inferences. It ultimately creates a multimodal skeleton for the representation of our distal and peri-personal space. Mediated by subcortical sensory systems and early primary sensory cortical areas, such processes remain complex and difficult to integrate in a unified model based on neural data. The perspective reviews the neuronal processes in primary sensory cortical areas known to underlie low-level perception and identify potential building blocks for any realistic model of early sensory processing in the brain of higher mammals.

By illustrating the complexity of explaining perceptual processes in terms of realistic neural-based architecture and rules, we try to identify bottleneck issues limiting presently further progresses:

- a purely bottom-up strategy on its own seems doomed to fail; conceptual approaches must be developed to reduce structural complexity.
- Multiple animal models are needed. Comparative studies show that primate brains are not simply inflated versions of rodent brains. Typical long-distance axons in rodents do not only remain within their cortical area of origin as in the ferret or the cat and rather tend to link multiple areas, sensory, limbic and motor. If long-range connections underlie our "perceptual grammar", rodents may actually have a very different perceptual language than higher mammals.
- If the choice of rodents as the reference model may be a deceiving alley for studying visual processing and more specifically the neural correlates of Gestalt laws, it may come at its advantage when searching for mechanisms responsible for olfaction, tactile sensing or for multimodal integration. The reduced size of the computational sheet makes that the higher visual cortical areas of the mouse abut directly other primary areas such as S1 and A1, with their interfacing border may constitute an ideal site for multimodal integration. In that respect, the mouse may offer interesting opportunities to understand the functional significance of heteromodal influences in primary areas, otherwise present but silent in primates.

We conclude that future progress in the field will depend on careful choices of experimental models (structures and species adapted to the questions under study), on the definition of agreed-upon naturalistic benchmarks for sensory stimulation, on the simultaneous acquisition of neural data on multiple spatio-temporal scales and on the identification of key principles (algorithms) supported both by data and simulations. These goals emphasize the inescapable roles of well-designed experiments, and on comparative approaches.

Neural correlates of unimodal perception and self-organization of internal knowledge in mammalian primary cortical areas

Task T3.5.1 - Yves Frégnac (CNRS-UNIC Unité de Neurosciences Information et Complexité)

Goals of the task

This task focuses on two issues:
- Sub-task 1: identifying common principles of sensory processing by cortical analyzers at various levels of integration, ranging from the conductance activation level (measured in vivo using both current clamp and voltage clamp techniques), spiking of single cells to the mesoscopic level (local field potentials, multiple recordings, EEG...);
- Sub-task 2: identifying neural processes in early visual cortical areas responsible for the emergence and self-organization of internal knowledge, tackling issues related to local vs. global feature processing, with an emphasis on apparent motion, motion extrapolation, prediction, grouping and completion.

Experimental datasets

- **Dataset sub-task 1: Multiscale study of reliability and correlation of evoked cortical dynamics during natural scene processing in cat primary visual cortex**

The principle of efficient coding suggests that visual processing in early sensory systems should be adapted to the statistical properties of the stimulus. By comparing intracellular responses to stimulus statistics of different complexity, we showed previously (Baudot *et al.*, 2013) that the temporal reliability of the neural code is optimized for natural statistics and that the stimulus-locked trial-to-trial variability of the subthreshold membrane potential waveforms is modulated by the statistics of the full field stimulus context.

Using the exact same stimulus seed, we performed a multiscale analysis based on more mesoscopic measures including multiple unit recordings (SUA and MUA) and local field potentials (LFP). The aim was to explore if the single-cell observations can be related (or not) to specific behavior and stimulus dependency shared by local ensemble of neurons and if a laminar dependency of the observed effects can be detected by these mesoscopic methods and what is the global impact of input statistics changes on the correlation between neurons.

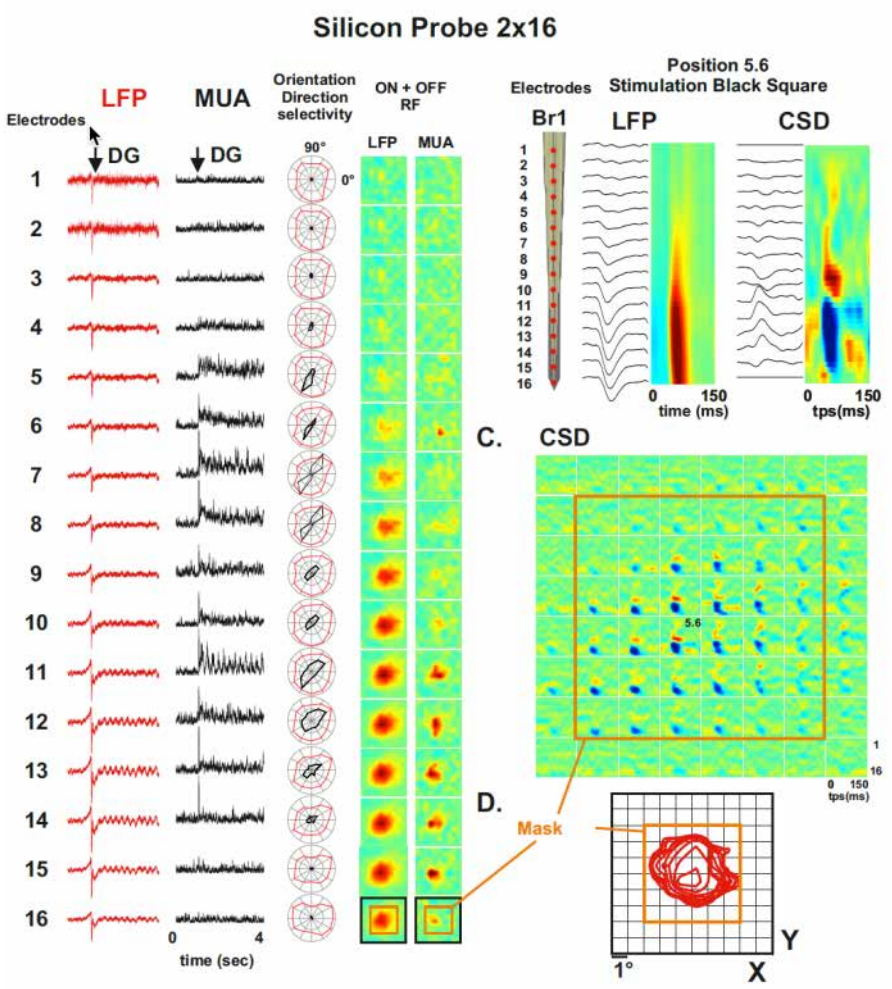


Figure 1: Example of data collection

Illustrating multiple recordings of local field potentials (LFP) and Multi-unit activity (MUA) with Michigan probes, with their associated functional measurements (Orientation/Direction tuning curves, ON-OFF receptive field maps) and Current source density (CSD) reconstructions.

Experimental design: In the area 17 of the anesthetized and paralyzed cat, we used Michigan silicon probes with different designs to realize laminar and lateral recordings across and within layers. To study the stimuli dependency of the reliability and correlation, we used the same stimulus benchmarks as in the Baudot et al study, i.e different types of visual stimuli with various statistic of increasing complexity (Figure 1): drifting gratings, gratings and natural Image animated with virtual eye-movements and dense noise stimuli. To ascertain the feedforward and local vs lateral nature of the field potentials, we partitioned the full field stimulation in a central mask large enough to cover the LFP RF (equivalent to the aggregate RF of the hypercolumn) and stimuli were presented in the three center/surround partitions (center only; surround only ; center + surround).

Results: For the LFP signal, the frequency content and its reliability (measured with coherence and wavelet analysis) were highly dependent of the type of stimuli and of the layer of the recordings. Similar conclusions were obtained for spiking activity with mean rate, sparseness, Fano-factor and noise correlation measures. In particular large

synchronizations of activity were found with natural image animated with saccade movement when the surround was stimulated (surround-only & center + surround).

This work has been presented at the Society for Neuroscience Meeting (Passarelli et al, 2015) and a Ms is in preparation.

❖ **Dataset sub-task 2: Evidence for a synaptic substrate of Gestalt Laws in cat V1**

The computational role of primary visual cortex (V1) in low-level perception remains largely debated. A dominant view assumes the prevalence of higher cortical areas and top-down processes in binding information across the visual field (Gilbert and Li, 2012). Long range "horizontal connectivity" in primary visual cortex (V1) has long been proposed to be the neural architecture substrate of "pop-out" perception, which does not require attention. However, this hypothesis relies exclusively on either anatomical correlates or indirect psychophysical data and has never been tested at the intracellular level. Here, we investigated the role of long-distance intracortical connections in form and motion processing by measuring, with intracellular recordings, their synaptic impact on neurons in area 17 (V1) of the anesthetized cat.

Experimental design and working hypothesis

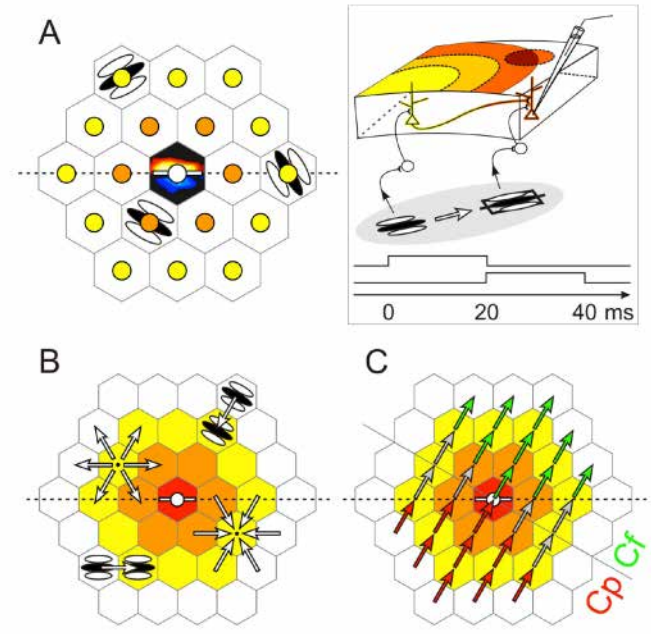


Figure 2: Stimulation protocol for probing spatial and axial motion sensitivity in the "silent surround of V1 receptive fields (see text).

Using an hexagonal stimulation node matrix (Figure 2A) centered on each recorded receptive field (RF, central black tile), we explored systematically the visual subthreshold (synaptic) responses in the "silent" (non-spiking) surround of V1 RFs to static stimuli (oriented Gabor patch), as well as to two-stroke apparent motion flow coaligned or cross-oriented with the orientation of the Gabor inducer (Figure 2B). The apparent motion sequence in visual space was matched in speed with that of horizontal propagation in cortical space (Bringuier et al, 1999; see top left inset in Figure 2).

In order to provide a meaningful estimate of the "horizontal" intracortical kernel, the analysis required the averaging of a couple dozen intracellularly recorded responses. This "mean-field" observation was necessary to overcome individual cell variability and reveal the "silent" synaptic footprint of the neural architecture needed to implement Gestalt laws.

Results

By systematically mapping synaptic responses to stimuli presented in the non-spiking surround of V1 receptive fields, we provide the first quantitative characterization of the lateral functional connectivity kernel of V1 neurons. Our results revealed at the population level two structuro-functional biases in the synaptic integration and dynamic association properties of V1 neurons.

First, subthreshold responses to oriented stimuli flashed in isolation in the non-spiking surround exhibited a geometrical organization around the preferred orientation axis mirroring the psychophysical "association field" for collinear contour perception (Hess, Hayes and Field, 1993)(Figure 2).

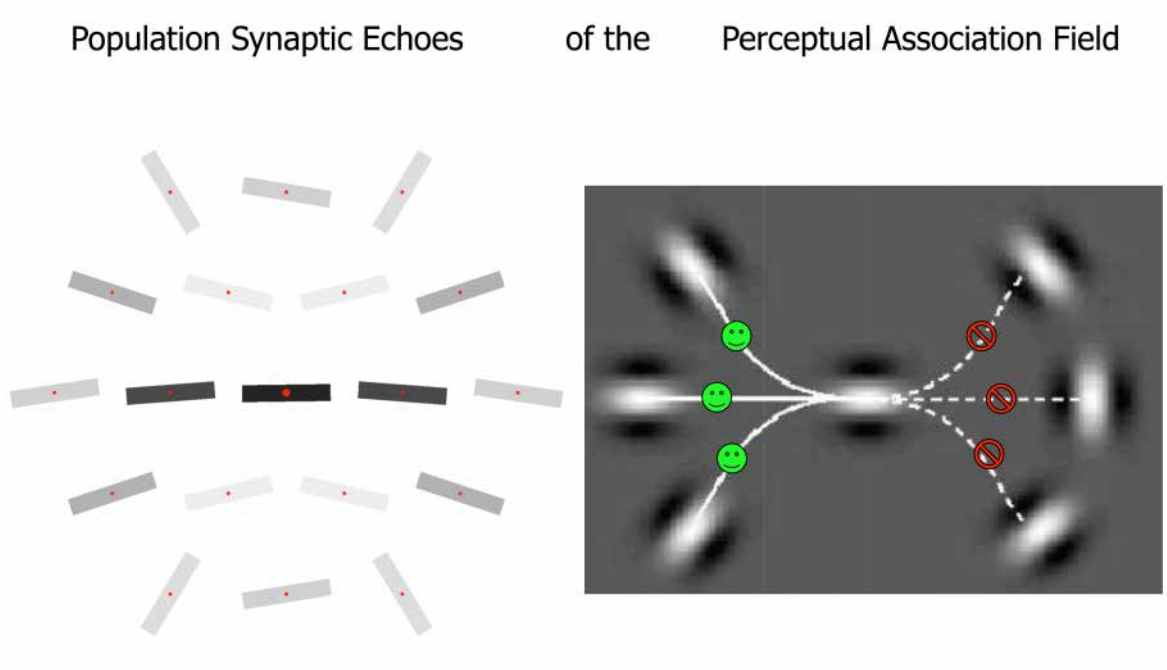


Figure 3: Synaptic correlate of the perceptual "Association Field" in V1.

Left panel, for each node in each cell, the preferred orientation of the synaptic responses is computed from the circular average of the individual responses to each orientation. The mean synaptic association field is obtained by performing a circular average over the cell population of the preferred orientations, at each node. The reliability of the orientation bias is represented by the level of grey of each bar. Right panel: Note the similarity in the pattern with the perceptual Association Field of Hess, Hayes and Field (1993).

Second, apparent motion stimuli, for which horizontal and feedforward synaptic inputs summed in-phase, evoked dominantly facilitatory non-linear interactions, specifically during centripetal collinear activation along the preferred orientation axis, at saccadic-like speeds. This spatio-temporal integration property suggests that local (orientation) and global (motion) information are already linked within V1. These electrophysiological results constitute the neural correlate of the preference shown by humans to perceive collinear sequences "faster" than "parallel" (Georges et al, 2002). This study is now in press in *The Journal of Neuroscience* (Gerard-Mercier et al, 2016).

Population synaptic Echoes of the Dynamic Association Field

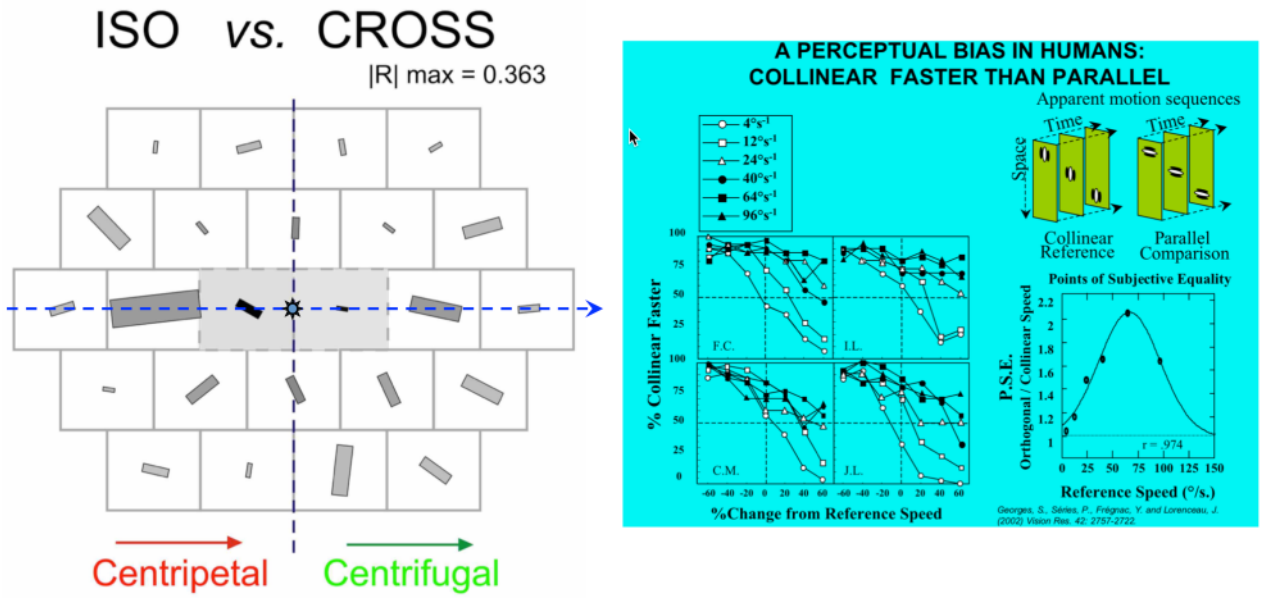


Figure 4: The dynamic association field concept (D-AF).

Left panel, the analysis requires to realign each RF map on the Apparent Motion axis (dotted horizontal arrow) and is based on the circular average of responses for common trajectories (defined by rotation invariance around the RF center). The resulting plot represents for each trajectory (centripetal, left half; centrifugal right half) the RF orientation maximizing the response to an axial AM flow, relative to the flow direction. The trajectories invading or leaving the RF (blue star) are represented as shaded rectangles. The comparison between responses for co-aligned motion (ISO) vs cross-oriented motion (CROSS) shows that apparent motion at saccadic speed make V1 neurons integrate co-aligned stimuli along their main axis instead of across the RF width axis (as classically reported at low speed). In other words, V1 neurons flip dynamically their axis of motion sensitivity by 90° when stimulated at saccadic speeds, retaining some capacity for broadband spatial integration for elongated contrast edges of the cell's preferred orientation

Right panel, this non-linear effect constitutes a neural correlate as early as V1 of a human psychophysical bias in motion flow detection: collinear sequences (ISO) are perceived as moving "faster" than cross-oriented (CROSS) configuration sequences (Georges et al, 2002).

Conclusion

We thus provide evidence for two neural correlates of low level perception, closely dependent on the spatiotemporal features of the synaptic integration field of V1 neurons, likely linked to intra-V1 horizontal connectivity. We suggest a new concept of dynamic association field, whose spatial anisotropy and extent are transiently updated and reconfigurated as a function of changes in the retinal flow statistics imposed during visuomotor exploration of natural scenes.

Data Provenance

The data were acquired by Yves FRÉGNAC's team at CNRS-UNIC (Unité de Neurosciences Information et Complexité), in Gif sur Yvette, in cats bred by the CNRS animal care facilities.

Provided data set

Due to the complexity of the in vivo preparation, such data cannot be exploited without intelligent understanding to the complete metadata. Since HBP and INCF platforms did not provide any specialized support to define some universally accepted format of metadata in vivo (which would require several person months work), we will open our own database to external request on the condition of a formal collaboration agreement relying on the expertise of the data provider.

Currently only a small subset of UNIC data is available online on a UNIC server, to which we have added a new portal subcomponent corresponding to the new data presented in this report: https://hbp.unic.cnrs-gif.fr/db

Note that the experimental costs of the HBP-related data were supported by the CNRS, the Agence Nationale de la Recherche and other international programs, which explains why the data web server will remain located at UNIC and under its control, while allowing yet-to-be defined collaborations with HBP users.

Note also that the cat experimental model is no longer included in the newly defined HBP objectives.

A Dataset Information Card has been completed (See DIC Task T3.5.1 "Recordings from primary visual cortex of anaesthetised cat during visual stimulation").

Collaborations

We collaborate closely with Andrew DAVISON (task leader in SP5 and SP9), with whom we co-supervise PhD students and Postdoctoral fellows (Jan ANTOLIK, Domenico GUARINO). We collaborate closely with Alain DESTEXHE (Director of SP4) and Olivier MARRE (SP4) on power-law analysis and correlation studies in asynchronous irregular networks. We collaborate with Olivier FAUGERAS (SP4) to promote links between Neuroscience and Mathematics through the organization of interdisciplinary conferences.

In collaboration with SP4, Yves FREGNAC organized two international symposia to be held at the EITN on "The Early Visual System", to better define the bottleneck issues that should be overcome to provide a realistic data-driven model of V1, adapting at multiple scales of integration to changes in sensory input statistics.

Neural correlates of unimodal and multi-modal perception in mammalian primary sensory areas

Task T3.5.2 - Brice Bathellier (CNRS-UNIC Unité de Neurosciences Information et Complexité)

Goal of the task

Sensory processing occurs throughout complex neural systems organized often described as a collection of separate feedforward pathways dedicated to each individual sensory modalities, which then merge in so called associative areas. However useful, this view is only a first approximation of the real structure of sensory system. First, associative areas send feedbacks projections to areas classically described as purely unisensory, thereby potentially channelling some crossmodal information. Second, sensory areas are even often directly interconnected with each other, although the purpose of these lateral connections is mainly unknown (Cappe et al., 2009). One may suppose however that all these crossmodal communication pathway serve to modulate or complement the information obtained with one modality based on information acquired with other sensory modalities (Ghazanfar and Schroeder, 2006). This role is actually suggested by multisensory illusions such as the ventriloquist effect in which a clear biasing of about the perceived location of a sound source occurs when a concomitant visual stimulus is given to a subject (Bonath et al., 2007). Symmetrically, in the double flash illusion, the presence of a double tone gives the erroneous impression that the visual stimulus is doubled in time (Apthorp et al., 2013).

So far however, the neural underpinning of such illusions is still elusive in most cases, although human imaging data suggests that primary sensory cortex is involved. Even worse, very little information exist about whether and how the areas classically described as unimodal represent information from other modalities.

To start filling this knowledge gap about the cognitive architecture involved in multisensory perception, the mouse model may be particularly useful. First, in the mouse, there has been novel high throughput anatomical studies establishing precise connectivity atlas between cortical areas and detailing multimodal connections. (Zingg et al., 2014) Second, new imaging tools such as two-photon imaging of genetically encoded calcium sensors permit in the mouse to measure sensory representation with cellular resolution in the awake animal at a throughput unmatched in any other animal model. The goal of this task actually only partially financed by HBP (only half-time postdoc salary), was to use these techniques to start extensively characterising the auditory visual representations in the mouse primary visual and auditory cortex.

Experimental design

To allow for repeated imaging of large neuronal populations in supragranular cortical layers, we surgically implanted chronic cranial windows over the auditory or visual cortex coupled with AAV virus injections to express the genetically encoded calcium indicator GCAMP6s. Mice were then head-fixed under the two-photon microscope and a set of auditory, visual and bimodal stimuli were played in different light conditions (darkness or screen illumination). Eye movements were monitored with a video camera, which proved very important for controlling this experiment as sounds can elicit eye movements. Stimulus delivery, data analysis and eye tracking were programmed with software developed in my team by Thomas Deneux (postdoc) and Alexandre Kempf (PhD student).

The stimuli consisted in a set of looming auditory and visual stimuli (increasing or decreasing in size or sound intensity) to assess potential auditory visual interactions

complemented with drifting gratings and frequency modulated sounds to assess typical tuning properties found in visual and auditory cortex. Looming stimuli were played either alone or in a bimodal conditions (all combinations were tested).

Experimental datasets

We have acquired a V1 dataset of 9056 neurons across 5 mice and an A1 data set of 3586 neurons across 7 mice and we have extensively analysed both calcium signals and eye movements (only during V1 recordings).

Eye movements

Eye tracking reveals that all sounds occasionally evoke eye movement in a time-locked manner (Figure 5). In the light these eye movements produce visual responses simply due to changes in the visual field. On the contrary, in the dark there is no visual response induced by sound induced eye movements (Figure 5).

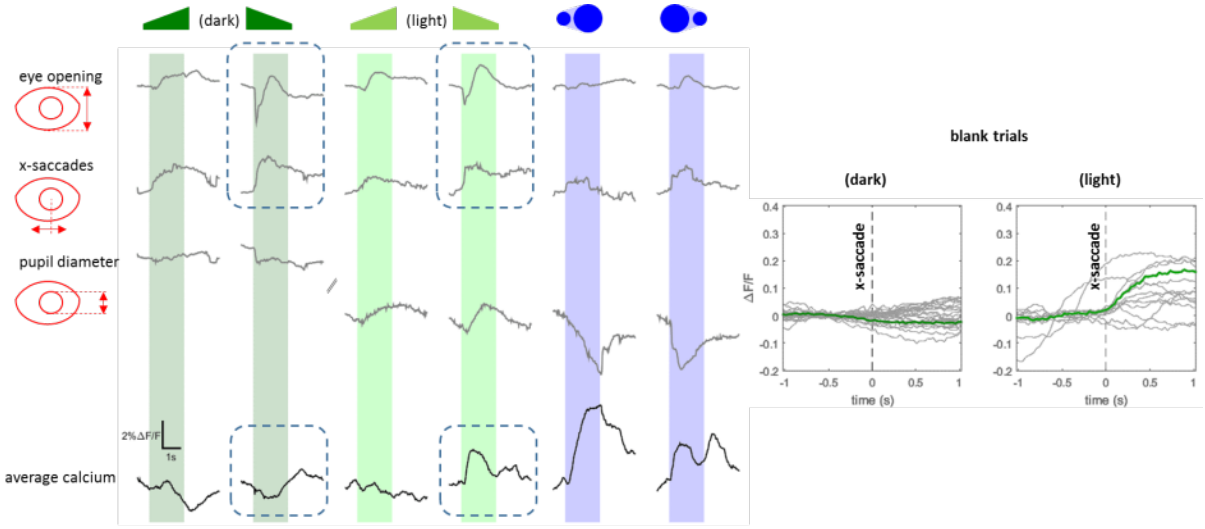


Figure 5

(left) Average eye tracking and calcium imaging signals for one recording session of in a mouse. The responses are shown for sounds in the dark or in the light or for visual stimuli. Sounds produce occasional eye responses that are more frequent at loud sound onsets. In the light, this results in a positive calcium signal corresponding to V1 responses due to changes in visual inputs evoked by eyes movements. These responses are absent in the dark. (right) Saccade-evoked visual responses are present in the light but absent in the dark

Thus we were led to conclude that potential auditory responses in V1 can only be revealed in awake animals if the sound induced eye movement effects are carefully compensated or in complete darkness.

Absence of visual responses in auditory cortex

We have extensively recorded region across the primary and secondary auditory cortex of the mouse and found no indication of statistically significant responses visually evoked responses in auditory areas. The analysis was done using generalized linear model

(parametric) or using non-parametric test and both methods converged to the same results. We also tested if the presence of a visual stimulus modulates auditory responses, using similar tests, and we found no effect. Hence, for the stimuli and technique used, we believe that there is no detectable functional visual input primary auditory cortex.

Negative and positive auditory responses in visual cortex

Analysing the responses to auditory stimulations in darkness we found that most neurons displayed a slight sound-evoked decrease in the calcium signal, which was evident at population level (Figure 5). This probably reflect a global inhibitory input from auditory cortex to V1 as shown previously with *in vivo* intracellular recordings (Iurilli et al., 2012). However, using non-parametric statistical test we also found that a small fraction of visual neurons (7.6% for a total population of 9056 cells tested with $p<.01$) responded positively to sounds in darkness. Because in this condition there is no visual input, these most likely correspond to auditory inputs. We were reinforced in this idea when observing a diversity of responses in these positively responding neurons, indicative of tuning to diverse sound parameters (Figure 6).

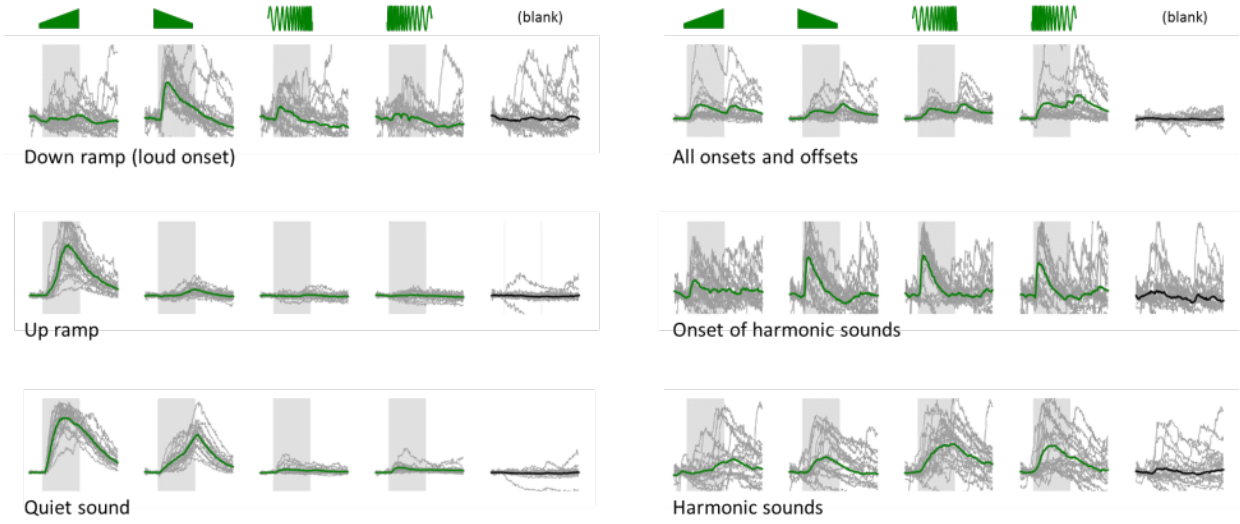


Figure 6: Examples of diverse auditory responses for 6 neurons in the visual cortex (in darkness).

Some neurons preferred the most quiet part of the looming sounds some preferred the loud onsets. Some preferred responded to frequency ramps and some did not. All these complex features are actually found in the auditory cortex to a great abundancy. This suggests direct connections between visual and auditory cortex as shown already for the negative subthreshold responses (Iurilli et al., 2012) (note that positive responses were not detected previously).

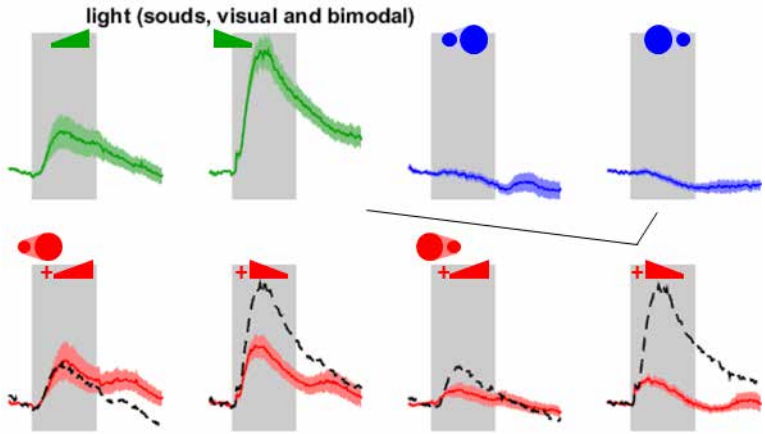


Figure 7: Example of nonlinear responses in bimodal condition for one neuron. Green: auditory stimulation alone.

Blue: Visual stimulation alone. Red: Auditory-visual stimulation. Black line: predicted additive summation.

Bimodal responses

We then checked whether auditory responding neurons (positive responses only) are perturbed by eye movements, and found that actually they have the same responses to sounds in the dark and in the light. Interestingly, these neurons can also respond visual stimuli and in most cases auditory and visual responses sum additively (not shown). But we found that some neurons also respond in a nonlinear manner signalling particular combinations of auditory and visual stimuli (Figure 7).

Conclusions

This project leads us to conclude that in the mouse there exist functional excitatory interactions leading to a sparse representation of sound features in the visual encoding space. This representation adds up mostly linearly to the visual representation but some even sparser nonlinear cells might encode specific combination of auditory and visual features.

Completeness of the dataset

Through this project, we have acquired data from more than 9000 neurons in visual cortex and more than 3000 in auditory cortex. This represents very high sample sizes sufficient to demonstrate with clear statistical significance the presence or absence of multisensory interactions in these areas.

Data Provenance

The data were acquired by Brice BATHELLIER's team at CNRS-UNIC (Unité de Neurosciences Information et Complexité) in 2014. The data is taken from awake Bl6C57-J mice.

Provided data set

We will provide the datasets for this study, ideally on the HBP collaborative servers. We have submitted a Dataset Information Card (see DIC Task T3.5.2 “Multimodal activity in visual cortex V1 of the mouse”), and the data is stored at:

http://sp3.s3.data.kit.edu/3.5.2/Auditory_visual_dataset/

The data set precisely consists in two-photon calcium imaging of mouse V1 and A1 activity during time-varying auditory visual stimulation. It is available in Matlab format and is otherwise used for a publication in preparation.

Collaborations

We are presently collaborating with Wolfgang MAASS (SP4) to analyse the two-photon calcium imaging data in the mouse auditory cortex, that was acquired by Brice BATHELLIER.

Annex: Publications related to WP 3.5

1. Interactions with other SPs :

- Interactions of UNIC experimenters have been occurring continuously during the Ramp-Up Phase with SP4 members (Alain Destexhe, Gustavo Decco, Viktor Jirsa, Olivier Marre) and with SP6 members (Andrew Davison), in the context of the European Institute of Theoretical Neuroscience and internal meeting at UNIC (Gif-sur-Yvette).
- Data-driven modelling of V1 done by Frégnac's team will be pursued with Jan Antolik (HBP PostDoc) and Andrew Davison (CNRS-UNIC; SP6) during phase 2 of HBP.

2. Organization of international conferences :

- Organization of an international workshop "The Early Visual System", held at the European Institute of Theoretical Neuroscience (EITN), Paris on January 19th-20th 2016, Paris.

  Organizers: Olivier Marre (IDV), Shilom Ullman (Weizmann), Yves Frégnac (CNRS UNIC), Alain Destexhe (CNRS UNIC)

  http://eitnconf-190116.sciencesconf.org/

- A follow-up two-day meeting is planned on May 19th and 20th, also to be held at the EITN (Paris), with the cofinancing of the Lidex Saclay I-Code (Y. Fregnac, SP3) and HBP-SP4 (A. Destexhe), gathering leading experts of the visual cortex in rodents and mammals, and confronting experimental and theoretical approaches.

3. Publications accepted or in press in refereed Journals or edited Books:

- **Frégnac, Y. and Bathellier, B. (2015).** Cortical correlates of low-level perception : from neural circuits to percepts. **Neuron 88**(1): 110-126.
- **Gerard-Mercier, F., Pananceau, M., Carelli, P., Troncoso, X. and Frégnac, Y.** (in press). Synaptic correlates of low-level perception in V1. **The Journal of Neuroscience**.
- **Frégnac, Y., Fournier, J., Gérard-Mercier, F., Monier, C., Pananceau, M., Carelli, P. and Troncoso, X. (2015).** The Visual Brain: computing through multiscale complexity. In "Micro-, Meso- and Macro-Dynamics of the Brain". **Research and Perspectives in Neurosciences**, Eds G. Buszaki and Y. Christen. Springer-Verlag, DOI 10.1007/978-3-319-28802-4_4

4. Experimental Work partly supported by HBP presented at the Society for Neuroscience:

- **Troncoso, X., Pananceau, M., LeBec, B., Desbois, C., Gerard-Mercier, F. and Y. Frégnac (2015).** Spatio-temporal synergy requirements for binding feedforward and horizontal waves in V1. **American Society for Neuroscience.** 331.04, P42, Chicago, USA.

- **Passarrelli, Y ., Foubert, L., Frégnac, Y. and Monier, C. (2015).** Multiscale study of reliability and correlation of evoked cortical dynamics during natural scène processing in cat primary visula cortex. **American Society for Neuroscience**. 331.20, Q16, Chicago, USA.
- **Deneux, D. Kempf, A. and Bathellier B. (2015).** Non-symmetric auditory-visual interactions at perceptual and cortical levels in mice. **American Society for Neuroscience**. 509.08, N21, Chicago, USA.
- **Deneux, D., Kempf, A., Ponsot, E. and Bathellier B. (2015).** Cortical population nonlinearities reflect asymmetric auditory perception in mice. **American Society for Neuroscience**. 231.05, J33, Chicago, USA.

5. Invited or plenary talks at International Conferences:

- **Frégnac, Y.** (2014). The Visual Brain : Computing with Complexity. **Heller Lecture Series in Computational Neuroscience**. Edmond and Lily Safra Center for Brain Sciences. The Hebrew University of Jerusalem. Israel (June 10, 2014). Plenary Lecture.

  http://elsc.huji.ac.il/content/heller-lecture-yves-frégnac
- **Frégnac, Y. (2014).** Hidden complexity in visual cortical receptive fields. **ELSC Seminar Series in Computational Neuroscience**. The Hebrew University of Jerusalem. Israel. Invited Conference.

  http://elsc.huji.ac.il/content/elsc-seminar-yves-fregnac
- **Frégnac, Y.** (2014). Perceptual association waves and collective belief in visual cortex. In Workshop « Geometrical models in Vision ». Org. F. Chittaro, G. Citti, JP Gauthier, A. Sarti and J. Petitot. **Trimester in « Geometry, Analysis and Dynamics on subRiemannian manifolds ».** Institut Henry Poincaré. Invited Conference.

  http://www.cmap.polytechnique.fr/subriemannian/ http://gmvision.lsis.org/
- **Bathellier B, (2014)** Some thoughts about auditory coding, perception and learning in mice. Bernstein Workshop "Population Codes: From Data Analysis to Mechanisms", Tutzing, Germany.
- **Frégnac, Y**. (2015). « What computational principles in artificial vision can be learnt from the biology of natural vision ». In the International Symposium « **From Neurons to Robots** ». Celebration of the **Silver Anniversary of the Interdisciplinary France-Uruguay**, Conférence en l'honneur du Dr Kirsty Grant, Montevideo, Uruguay. Invited Conference.
- **Frégnac, Y.** (2015). The Visual Brain : Computing through multiscale complexity. **1st International Conference on Mathematical Neuroscience (ICNMS).** Org. O. Faaugeras. Juan les Pins, INRIA (Plenary Lecture).

  https://icmns2015.inria.fr/files/2015/06/E_Program_Officiel_ICMNS2015.pdf
- **Frégnac, Y., Sarti, A. and Antolik, J. (**2015). Which theory to describe V1? In **"Confronting mean-field theories to measurements: a perspective from Neuroscience**". Orgs Cessac, B. and Faugeras, O., **BrainScales Symposium at EITN**, Paris. Invited Conference.

- **Frégnac, Y.** (2015). The visual Brain: computing through multiscale complexity. In **"Micro-, meso- and macro-dynamics of the Brain"**. Orgs. Buszaki, G. and Christen, Y. Conference Ipsen, Paris.
- **Frégnac, Y.** (2015). Context-dependent adaptive computing in the early visual system. **3rd Canonical Neural Computation Conference**, Orgs. M. Carandini, D. Heeger and T. Movshon, New York University, Firenze, Italy. (Invited Conference).
- **Bathellier B,** (2015) Cortical correlates of asymmetric auditory perception in mice. NYU Abu Dhabi Workshop on Computational and Experimental Neuroscience, Abu Dhabi, UAE.
- **Bathellier B,** (2015) Cortical population nonlinearities explain asymmetric auditory perception. Annual meeting of Neuroscience School Paris (ENP), La Clusaz, France.
- **Bathellier B,** (2015) Complex features in uni- and multi-modal representations of primary sensory cortex **Annual SP3 conference**, Foundation Hugot, College de France, Paris, France.

National and International tribunes

- **Frégnac, Y. (2013).** Big science needs big concepts. In Voices: BRAIN Initiative and Human Brain Project: Hopes and Reservations **Cell**, **155**(2): 265–266.
- **Frégnac, Y.** (2014). A CNRS view of the Human Brain Project. Réunion sur « **The Human Brain Project** » organisée par la Direction des Relations internationales du CNRS. Meudon (24 Février 2014), **CNRS**. Invited Conference.
- **Frégnac, Y.** (2014). A CNRS view of the Human Brain Project. Réunion sur « **The Human Brain Project** » organisée par la Direction des Relations internationales du CNRS. Meudon (24 Février 2014), **CNRS**. Invited Conference.
- **Frégnac, Y. and Laurent, G.** (2014). Where is the brain in the Human Brain Project. Nature Comments. **Nature, 513** : 27-29. Supplementary to be found at www.brain.mpg.de/fregnac-laurent-2014
- **Frégnac, Y.** (2015). Human Brain Project : « La médiation laisse espérer que la crise sera réversible. » *Pour la Science* : Décember issue.

  http://www.pourlascience.fr/ewb_pages/a/actu-human-brain-project-nouvelles-orientations-36172.php
- **Frégnac, Y.**: Participation to the report at the CSH-Asia Conference on Big Data projects and Brain Sciences:

  Contribution included in **Huang, J.Z. and Luo, L. (2015).** Perspective : Neuroscience : It takes the world to understand the brain. **Science :** 350 (6256) : 42-44.

General audience conferences

Frégnac, Y. and Grollier, S. (2014). Vers un Cerveau simulé par Ordinateur. Cycle de Conférences « **Dialogues – des clés pour comprendre** » organisé par le CNRS (INS2I). **Musée des Arts et Métiers**. Conférence Grand Public.

Frégnac, Y. (2015). Interdisciplinarité : du cerveau biologique au cerveau virtuel . Cycle de Conférences « le cerveau virtuel est-il bien équilibré ? » organisé par Mahfoud Chkouri. **Cité des Sciences et de l'Industrie**. Conférence Grand Public.

1.2 Electrophysiological signals from early visual cortex: data and model

Task T3.1.4 - Peter De Weerd (UM), Mark Roberts (UM), Avgis Hadjipapas (UNIC), Margarita Zachariou (UNIC)

Overview

We aimed to create a model of V1 gamma generating networks which would serve to link the observations of human V1 gamma measured using MEG with that of gamma (single unit spikes and LFPs) measured invasively in macaque V1 (Hadjipapas et al., 2015). Our central hypothesis was that differences between these signals arose principally from the differences in the spatial scale of the measurement; whereby invasive signals represent the behavior of single units (micro-scale) or local networks (LFP, meso-scale) while the MEG represents the aggregation of large populations of neurons in the vertical and horizontal dimension (macro-scale). The MEG is thereby sensitive to synchronization and time delays between local networks as well as the behavior within local networks. To most important step in achieving this is to create a simplified, abstract model with a sufficiently restricted space of relevant parameters that can be heavily constrained by empirical measurements from macaque V1. The cortical sheet is made of multiple cortical columns, which are linked by horizontal connectivity spreading across upper layers. Within each column neuronal networks operate in separate layers and are linked by vertical connectivity originating from excitatory neurons in deep layers. We examined the activity of this large network in a series of steps: We first developed an undifferentiated model constrained by observations of the contrast response function of 1) single unit spike rates 2) gamma power 3) gamma peak frequency. To expand this model to include laminar compartments and horizontal connectivity across columns we examined 1) laminar differences in the empirical constraints 2) temporal dependencies between layers and between columns which will later be used as to constrain how model networks are combined to form a large scale model.

How to model a human brain?
Start in V1

Hadjipapas A, Roberts M, Zachariou M, Lowet E, De Weerd P.

Abstract

The ultimate outcome of the research in the combined disciplines of neuroscience and biology would be that it enabled us to model a human brain. Assuming that this would be possible, what is the best way to get there? In the present article, we first review a number of possible modelling approaches of varying scale and complexity. Based on that analysis, we suggest that efficient modelling requires integration with empirical validation with empirically -observed signals appropriate for the scale of the model. We suggest that the upscaling in size and complexity should be carried out in a manner that remains tightly linked with empirical validation using signals of an increased level of aggregation. Without claiming that oscillatory behaviour of a network is the only functional property of interest, we do suggest that oscillatory signals at different levels of aggregation can be valuable in empirically constraining parameter spaces of models at micro-, meso-, and macro-scales. We illustrate this by showing how a new model of gamma, built upon basic properties of early visual cortex, creates new insight into the fundamental properties by which early visual cortex operates, and how this new model reconciles seemingly irreconcilable data. We suggest that these mathematically well-specified models of gamma are based on local network structures that can be used as the building blocks for larger-scale models of early visual cortex.

1. Introduction

The goal of the present article is to review a selection of empirical studies that are useful for modelling primate early visual cortex, and to consider the positive and negative aspects of a number of modelling approaches. As such, this review may contribute to building a model of the human brain, as is the ambition of the Human Brain Project (HBP). It is useful to begin by situating the scope of our review within the overall scope of the HPB's ambitions. Modelling a human brain would require a full understanding of roughly 100 billion neurons in terms of their genomic and molecular processes and connectivity (each neuron connecting with ~10000 others). A complete human brain model would also include the interactions with glia cells, comprising about 50% of the brain's mass and performing a plethora of essential functions. Importantly, communication in the brain is not limited to neurons, but includes inter-glial communication, and neuronal-glial interactions determining sensory functions such as neuronal tuning (Perea et al., 2014) as well as neuronal plasticity (Fields, 2008). The different types of neurons (and glia) communicate with each other by a number of neurotransmitter and neuromodulator systems. All of these interactions in each adult human brain result, in part, from an individual's genome reflecting 2-3 million years of human evolution preceded by many more millions of years of primate evolution (Reed and Bidner, 2004), and from the individually experienced statistics of environmental stimulation throughout life and especially during early development. With our current limited understanding, a model of the scope of a human brain will be vastly under-constrained due to missing data and the resulting inability to empirically inform the large numbers of important and diverse parameters that would need to be set. A whole-brain model will therefore most likely show a host of unexpected behaviours that will require reduction to tractable problems to permit experimental and computational investigation. An alternative approach therefore would be to limit a model's overall scope and nature by the specific scientific goal to which the model is applied.

In the present review, we will focus on neuroscientific models of visual cortex, primarily the primary and secondary visual cortex (V1 and V2), because arguably there are no other two areas in the brain that are better understood. Despite the enormous body of knowledge that exists about V1 and V2, the same questions about the scope of modelling that exist for the brain as a whole, also exist for these early visual areas. V1 and V2 are complex regions of the brain characterized by cells showing a range of sensitivities to a range of stimulus parameters (e.g. (Hubel and Wiesel, 1959, 1960, 1962)), by exquisite laminar and columnar structure (Douglas and Martin, 2004; Stettler et al., 2002), and by highly specific thalamic as well as inter-areal connectivity (Sincich and Horton, 2005). Beyond spiking responses measured from single or multi-neurons, oscillatory behaviour as measured in population measures of neural activity is present in a range of frequency bands in both areas. An important source of oscillatory neural activity is thought to be the interaction between excitatory and inhibitory neurons, which is also of paramount importance for both the sensory tuning of neurons, and their spatial summation properties. Hence, fully understanding a range of well-established functional properties of V1 requires a very complete insight into the excitatory-inhibitory interactions. Reaching full or even sufficient insight is rendered difficult by the fact that excitatory and especially inhibitory neurons come in a great range of varieties (Buzsáki et al., 2004; Markram et al., 2004). In addition, their within-class interactions are still poorly understood, which further renders a full understanding of even the most ubiquitous properties in V1 and V2 difficult. For example, it has been well established that parvalbumin positive (PV+) interneurons, presumably basket-cells targeting the axosomatic region of pyramidal cells (Buzsáki and Wang, 2012; Traub et al., 1996), are critical for gamma-generation (Buzsáki and Wang, 2012; Cardin et al., 2009). However, somatostatin positive interneurons (SOM+), targeting dendritic regions of pyramidal cells and critical for surround suppression (Cottam et al., 2013) also have been proposed to play a role in cortical gamma oscillations (Gieselmann and Thiele, 2008). Recent findings have begun to elucidate the interactions among these two inhibitory cell types. SOM+ neurons have been shown to strongly modulate PV+ interneuron activity (Cottam et al., 2013), but knock-out does not eliminate gamma (Kuki et al., 2015). This suggests different roles of the two inhibitory cell types in generating and controlling gamma, but clearly more research on their interactions is required (for more on empirical and modelling studied of gamma, see Section 5). As a further challenge to modelling early visual cortex, even the more basic aspects of V1 such as the function of columns (Sincich and Horton, 2005) or the specific purpose and organization of connectivity between V1 and V2 (Sincich and Horton, 2005) are a matter of debate. This shows how formidable the challenge is to even 'just' model the early visual cortex. In the following section, different modelling approaches that can be useful in enhancing our understanding of visual cortex are presented.

2. Different categories of models

Within the field of (visual) neuroscience, there exist three broad subcategories of models depending on the level and kind of insight one wishes to achieve (i.e. the goal of the model). We will refer to them as (1) ***psychophysiological,*** **(2)** ***functional / computational,*** and (**3**) ***biophysical / generative.*** These broad categories of models aim for different levels of insight, and hence they use different theoretical concepts.

1) ***Psychophysiological models:*** These are conceptual models based on the observed correlation between a mental function, as studied in a cognitive or behavioural paradigm, and a measure of brain activity, such as for example BOLD in an fMRI study, or stimulus or task-related change in electrical potentials or in brain rhythms (oscillations) in a neurophysiological study. Their primary aim is to identify a *statistically (and experimentally) reliable* correlation between a well-specified cognitive

function and a brain correlate, and to use this correlational data to formulate a conceptual theory of how a set of brain areas or specific neurons in brain areas contribute to the studied cognitive function (Kiorpes et al., 2013; Teller, 1984).

2) ***Functional / computational models:*** This approach deals with specifying a mechanism that in principle *could* underlie a specific cognitive function. In the model implementation, the primary goal is to engineer a neural mechanism capable of performing a specific type of information processing underlying a cognitive function of interest. There is usually no great concern regarding the incorporation of realistic spatial scale or biophysical detail of model neurons. Instead, the models provide proof of concept of the potential relevance for a specific type of cognition or behaviour of a mechanism reduced to its essentials.

3) ***Biophysical / generative:*** This approach is concerned with providing a principled and realistic interpretation of the biophysical generative mechanisms underlying empirically measured signals (Einevoll et al., 2013). In these models, realistic neural networks are based on known biophysical principles of neurons, complemented with knowledge from cellular biology, anatomy and physiology. The primary aim in *generative models* is to achieve a model that is realistic enough to yield an understanding of the empirically measured signal. This refers to the crucial concept of *observability* (Kalman, 1959, 1963). Generative models often require a detailed specification of the so-called *measurement or observation functions.* This means that these models require a specification of the connection between the measured signals on the one hand, and the system variables or parameters of interest that are not directly observable on the other hand. These non-observable variables are also referred to as *hidden or latent* variables (Dodge, 2006; Tarantola, 2005). The key distinction between hidden and observable system variables originates from statistical estimation theory (Bollen, 2002; Kaplan, 2009) and control theory (Kalman, 1959, 1963; Kalman and Bucy, 1961; Kalman and others, 1960; Tarantola, 2005). To understand the hidden versus observable dichotomy, consider the following example: In the cortex a transmembrane current at a particular location of interest such as the apical dendrites of layer 5 pyramidal neurons is an important variable serving as a functionally-important input, yet it is unobservable. A biophysical model may be developed to extract this interesting unobservable variable from a directly measured *observable,* for instance extracellular LFPs measured at fixed locations of a laminar probe. The so-called *observation function* connects what we would like to study (the unobservable system variable*)* to what can be experimentally measured (the *observable*). Establishing the observation function corresponds to solving the *forward problem.* If the solution to the forward problem is sufficiently specified, then one can also attempt to solve the so-called *inverse problem* (Tarantola, 2005), by utilizing certain assumptions about the system under study. This generally refers to identifying the *most likely* underlying unobserved system variable or mechanism of interest (e.g., layer-specific transmembrane current) *given a specific set of measured observables* (e.g., extracellular LFPs from laminar probes).

Functional and biophysical models have a different relationship to empirical validation. Functional models are usually relatively abstract, and do not provide a sufficiently specified forward model, so that a quantitative or principled comparison with empirical data will be difficult. Thus, while theoretically valuable, such models may not be implemented at all in the brain. By contrast, detailed empirical validation is possible for biophysical models, as a sufficiently specified forward model permits investigating the crucial inverse question, that is, which generative mechanism is most likely given observed data. Hence for the more abstract, functional models to become amenable to empirical validation, a limited but sufficient set of biophysical detail must be incorporated, depending on the signal of interest and scientific question.

3. Selecting the appropriate kind of model given the scope of the research question

There are in essence two guiding principles for the construction of models. The first is structural /generative. Here, the goal is to achieve models in which substantial detail is incorporated, relevant for the generation of the observable output signals of interest (e.g. single unit spike trains in different layers, LFPs at different depths). Hence, these models incorporate a high degree of biophysical realism based on structural/anatomical knowledge of first principles. The basic idea is that if one achieves a model system with sufficient realism, it will show functionality in terms of the observable output signals similar to the various observed functions in its biological counterpart. Within this class of models, the primary aim is to build as realistic as possible single neurons, columns, or an entire cortical area. Given computational limitations there is a trade-off between the number of model neurons in a network and the biophysical realism of the single neuron models used. The second guiding principle is functional. Here, the scale, scope and complexity is determined by the functional behaviour that one wishes to model, and the model's structure and realism is adapted to the specific phenomenon one wishes to study.

3.1 Models aiming for structural realism of the data generating process

Before discussing models aiming for structural realism in modelling visual cortex, it is worth pointing out a few of its major features. In mammalian neocortex, substantial vertical, and horizontal structure exists (Binzegger et al., 2004; Douglas and Martin, 2004). The vertical structure entails interconnected cortical layers receiving differential inputs from the afferent visual

pathway. The horizontal structure is related to the interconnections among cortical-columns, which are layer dependent. Realistic cortical models must take into account the wide variety of cell types that exist in V1, as defined functionally and morphologically. Not only is there diversity in morphology of excitatory neurons, especially the diversity of inhibitory neurons is particularly striking (Buzsáki et al., 2004; Markram et al., 2004). Thus, a cortical area can be considered from a ***microscopic*** scale (neurons with specific morphologies within a certain layer or stretching across layers), a ***mesoscopic*** scale (networks of neurons within a layer, or crossing layers to form a *cortical column*) and ***macroscopic structure*** (an area itself ensuing from the interconnections between cortical columns). At each scale, different experimental observables can be obtained. For instance from laminar recordings, single unit spike trains (micro) and LFPs (meso) can be obtained. ECoG, MEG, and EEG on the other hand measure phenomena at the macroscopic scale.

To understand how one type of signal is related to another, cross-scale links must be made. This can be done experimentally, using simultaneous recordings at various scales (Musall et al., 2014) or using data at many scales acquired in the same experimental paradigm (Hadjipapas et al., 2015). Increased insight into cross-scale links can be achieved with generative models that incorporate vertical and horizontal structure and realistic cell types and morphology (Figure 8).

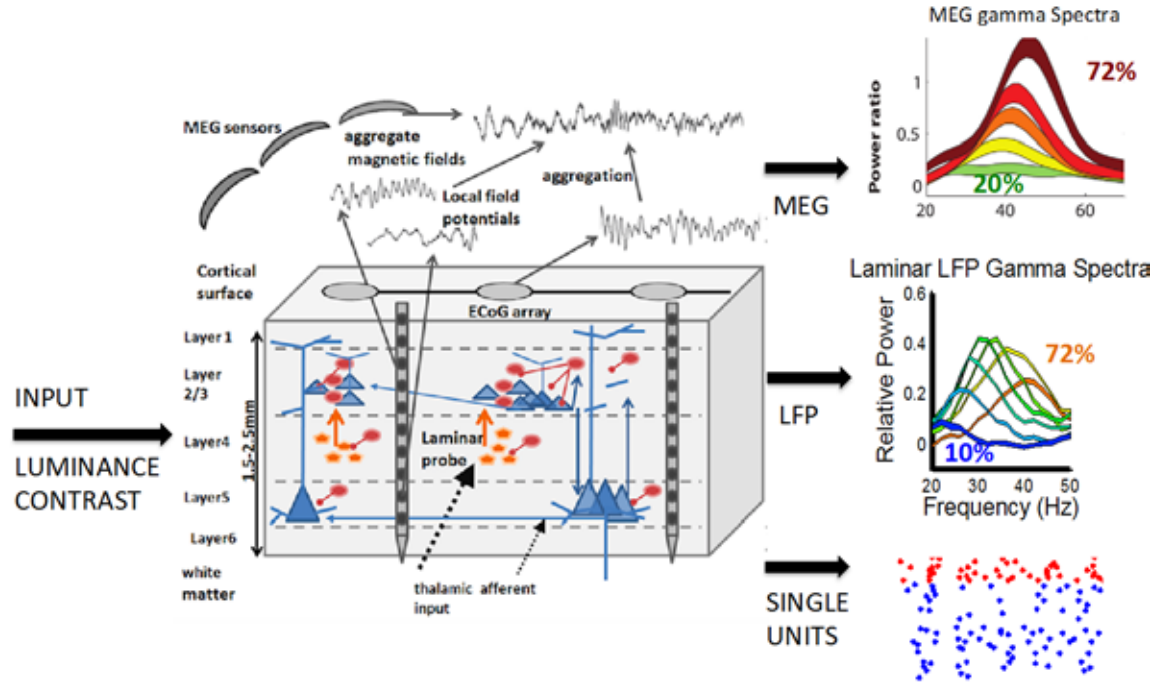


Figure 8: Schematic of primate primary visual cortex (V1) and changes in observable signals in response to experimental manipulation.

This figure schematically illustrates the approach taken in principled comparisons of empirical signals measured at different scales as in (Hadjipapas et al., 2015) *and modelled in* (Zachariou et al., 2015)*. In this highly simplified schematic, primary visual cortex consists of various types of interconnected neurons, of predominantly two functional classes, excitatory (blue) and inhibitory (red) (also illustrated in orange are layer-4 stellate cells). These are structured in six layers, whereby some neurons, especially large pyramidal neurons with cell body in layer 5 stretch across layers. There are connections between neurons both within and across layers. This vertical structure (interconnected layers forming a* ***cortical column****) is centred around these large pyramids. Cortical columns ensuing from this vertical structure, are replicated laterally but remain mutually-interconnected via direct horizontal and other (including indirect) connections. In experiments in nonhuman primate, e.g.* (Roberts et al., 2013)*, laminar probes were inserted. From these laminar probes, LFPs, CSDs* (Roberts et al., 2013) *and single unit spike trains were extracted in response to contrast-varied gratings. In humans the same stimulus paradigm was applied* (Hadjipapas et al., 2015)*, and source reconstructed MEG signal localized to human primary visual cortex was obtained. Increases in contrast caused typical shifts in gamma spectra towards higher frequencies. Other macroscopic observables, that could be theoretically obtained are subdural ECoG recordings and scalp EEG. A full model of V1 would involve incorporating all known anatomical and functional details, which can then be hoped to simulate a large array of functional phenomena of interest. Alternatively, the empirical data in a contrast manipulation paradigm could be used to progressively build up and constrain realistic neuronal and network behaviour at increasing scales* (Hadjipapas et al., 2015; Zachariou et al., 2015)*.*

A recent study provided a great example of a generative model achieving realism in the model-scale (Potjans and Diesmann, 2014). In this study, a cortical column model was constructed comprising almost 100,000 neurons. In addition, substantial realism in network structure and connectivity was employed. The model employs a realistic laminar structure in which crucial anatomical statistics such as the numbers and density of excitatory and inhibitory neurons is respected. In addition, the within and cross-laminar connectivities are also realistic, as they were directly derived from quantitative anatomical (Binzegger et al., 2004) and physiological data (Thomson et al., 2002). However, this high level of realism came at the cost of using relatively unrealistic integrate-and-fire neurons, rather than conductance-based Hodgkin and Huxley model neurons. These neurons mainly aim at reproducing realistic spike train dynamics, which implies that there is not much information on the membrane potential dynamics that underlie the single neuron and the aggregate signals such as the LFP. This in turn limits the extent to which this model can be validated with experimental LFP data. Nevertheless, the model produces spontaneous patterns of activity that are indeed realistic and reproduces key experimental observations with respect to the timing and duration of laminar evoked responses to brief stimuli (modelled as pulses in thalamic afferents). Importantly, this model was not constructed to replicate any of these functional phenomena. Instead, it was based on *first-principles* derived from known anatomy and physiology of the cortex. This exemplifies one of the key strengths of such large scale, realistic models, which is the ability to reproduce, at least in general terms, a large number of diverse dynamic repertoires.

A related recent study by the same group (van Albada et al., 2015) delivered a further important insight, namely that downscaling (i.e. going from the real situation of millions of neurons to thousands or hundreds) can affect network dynamics, even when classical corrections for downscaling are applied. This is because higher-order statistics such as neuronal correlations are not preserved by such corrections. In turn, this has a consequence for the observed dynamic behaviour of the downsized network, which is no longer the same as in the full-scale case. The work from this group has thus highlighted that to observe full network dynamics including neuronal correlations, one needs to consider a realistic number of neurons and synapses. This indicates an inherent limitation in much of the work performed in smaller scale models.

Whereas in one type of studies, simple spiking model neurons are used to model a column (van Albada et al., 2015; Potjans and Diesmann, 2014), a different generative approach focuses precisely on the generation of LFP from anatomically-detailed biophysical model neurons (Einevoll et al., 2013). Here, emphasis is placed on using realistic single model neurons, at the cost of a much lower number of model neurons. To enhance realism in model-neurons, they are built from a large number of compartments. Due to their realistic morphology, these model neurons and their compartments inherently take a realistic position in the laminar structure of the overall network in which they are embedded. For example, a pyramidal cell with its cell body in layer 5 will have a dendritic tree spanning the superficial layers, and axons in the deep layers. Because functional connectivity (inputs/outputs) is layer-specific, this will contribute to functionally different compartments. Hence, multi-compartmental neurons allow for the consideration of the distribution of synaptic inputs across the different cell compartments. The distribution of activity in the different compartments of model pyramidal cells determines the LFP signal properties, in a way that is highly analogous to what would be measured with a laminar probe experimentally. This model permits linking hidden variables (synaptic currents) to aggregate signals measured empirically, leading to a well-specified forward model. This approach thus has been fruitful in elucidating the nature and fundamental properties of the LFP signal such as the origin of the power law observed in the LFP (and EEG/MEG) power spectrum (Pettersen et al., 2014), the factors governing LFP spatial reach and power (Lindén et al., 2011), and the effects of signal frequency on LFP spatial reach (Łęski et al., 2013). Knowledge of these properties of LFP is also highly relevant for empirical studies examining LFP synchronization /coherence. Note that the approach taken by Einevol et al. (Einevoll et al., 2013; Łęski et al., 2013) was to start from first principles of cell function and morphology, functional anatomy, and solutions to well-defined bioelectric forward problems and volume conductor theory. Such models yield proof of principle as they reproduce many generic features of the measured LFP. The approach is generic, as there was no aim to reproduce LFP signals measured in any particular experimental situation. The ground-truth data from these models can be tested against experimental data, but can also be used to validate computationally-cheaper network models and algorithmically simpler LFP signal proxies (see (Mazzoni et al., 2015)).

In modeling realistic multicompartmental models, an additional tradeoff must be made. The multi-compartmental neurons are modeled typically as 'passive'; they lack active conductances and other, non-synaptic contributions to the LFP that are known to exist (Buzsáki et al., 2012a). In addition, they do not produce spiking behavior. Spiking multi-compartmental models that are mutually-interacting can be simulated , but at the costs of a simplified compartmental structure and smaller spatial scale (number of models neurons) (Jones et al., 2007, 2009; Lee and Jones, 2013). In making this trade-off, it is useful to consider however that the linearity that underlies the passive conduction models also allows computations over a large scale, permitting an easier derivation of analytical expressions of aggregate signals such as the LFP. Such formalism is useful for formulating the fundamental properties of the aggregate LFP see (Einevoll et al., 2013; Łęski et al., 2013; Lindén et al., 2011).

In between large-scale models that come at the cost of simplified model neurons, and small-scale models with highly realistic single neurons, a hybrid approach is also possible. Only recently, a hybrid scheme was constructed (Hagen et al., 2015), whereby a large-scale

model (Potjans and Diesmann, 2014) was utilized to provide input to a number of passive, multi-compartmental morphologically realistic neurons representative for each cortical layer. From these realistic multicompartmental neurons, a realistic LFP can be 'measured' (Hagen et al., 2015). This combines the advantages of a large-scale model with realistic neuronal density, connectivity, and correlation structure, with the robust LFP forward model that can be derived thanks to the multi-compartmental neurons. The model shows realistic behavior in terms of its computed CSD and LFP. Moreover, a study of model behavior showed that to obtain a realistic LFP, the cross-correlations between single (model) neuron LFP contributions need to be preserved, and these in turn depend much on preserving the correlational structure of the input. This insight points to an inherent limitation arising from using small numbers of model neurons, in terms of obtaining an accurate forward model of an aggregate signal. A sufficiently large number of neurons is necessary, otherwise, in addition to the limitation of not achieving realistic dynamics with small number of neurons (see (van Albada et al., 2015), the actual estimation of the aggregate is inaccurate (because of inaccuracies in the observation function). A sufficiently large number of neurons is essential because the scale and correlational/synchronization structure is important in determining what components will be observable and /or will dominate the aggregate signals (see (Hadjipapas et al., 2009, 2015), but also the classic work of Nunez and colleagues (Nunez and Srinivasan, 2006a)).

The inherent limitations of these large-scale and/or complex-structured modelling approaches are largely intuitive. First, in large-scale models an enormous number of parameters must be set, and to do so in an empirically validated manner is a formidable challenge. Second, the storing and analysing of the output of such models requires enormous storage and processing capacity. Rapid progress is being made however to cope with such large-scale data (e.g. Elephant, http://neuralensemble.org/elephant/), a challenge that is also being addressed by the HBP. Third, even after particular analyses have been done, the data may be so complex that they may be difficult to interpret. Therefore, currently large, structurally realistic models built from first principles are predominantly used in providing proof of basic concepts. So far, however, such models have had limited success in explaining or modelling more specific experimentally observed phenomena. The latter is central to achieving mechanistic insight specific to observed experimental data.

3.2 Models tailored for specific experimental observations

In models constructed to investigate the basis of specific empirical data, a first approach can be to limit the model to the appropriate scale and complexity of network, and the complexity of units and synapses required to understand the phenomenon. For example, gamma oscillations are locally generated, and hence it can be argued that the basic mechanisms of gamma as measured in the LFP can be understood by a model network of limited size and complexity, and with simplified integrate-and-firing neurons. With these simplifications, a price may be paid in the possibility for empirical validation, and hence a careful balance needs to be strived for (further discussion in section 3.2.2). However, when it is aimed to build a model reproducing macroscopic signals such as MEG /EEG, preserving a laminar structure in a sufficiently expanded model network will be important (Jones et al., 2009; Lee and Jones, 2013). This is because the generation of MEG/EEG is dependent on translaminar current dipoles, which depend in turn on differential behaviour of dendritic compartments in superficial layers and somatic compartments in deep layers, for example in the case of layer 5 pyramidal neurons. Thus, transmembrane currents essentially initiate a process whereby differences in charge between these two compartments cause axial intracellular (so called *impressed or primary*) and extracellular (so called secondary or return) currents. The impressed/primary currents aggregate over many (pyramidal) neurons that have the same orientation (vertically to the cortical surface), and form the current dipoles that are the main generators of MEG and EEG (Einevoll et al., 2013; Hadjipapas et al., 2015; Hämäläinen et al., 1993; Jones et al., 2009; Nunez, 2006). Thus, if one aims to reproduce MEG signals or their current dipole generators (Jones et al., 2009; Lee and Jones, 2013), then preserving some laminar structure and the key compartments (dendrites vs. soma) is important. One pioneering conceptual aspect in the work by Jones and colleagues (Jones et al., 2009; Lee and Jones, 2013), is the attempt to link the model with observed empirical data and reproduce very specific experimental findings and features of the measured MEG signals. Inverse MEG source reconstructions were undertaken on the real data to obtain the underlying current dipoles. Then current dipole moments (source signal time series) were analysed in the time and frequency domain to extract salient and specific experimental observables. They then employed a mathematical model, specifying the forward model underlying the generation of current dipole moments. This thus allowed for a direct comparison of equivalent model data and empirically measured signals (and thus for empirical validation of the model). One limitation in the interpretation of these studies is that aggregate signals from a small set of neurons may not be fully realistic (Hagen et al., 2015). Moreover, to use small-scale models to explain macroscopic measurements (MEG) also may represent a limitation, and empirical validation would have to include measurements at a smaller scale (e.g., spikes and LFPs from laminar recordings).

Below, we will focus on models of gamma oscillations, which in a number of theories, are essential for information transmission in the brain (see Section 5). Functional and generative models are considered, but we now focus on the emergence of gamma in visual cortex in response to visual stimulation.

3.2.1 ***Functional/computational models of gamma***

Such models address the aims of the functional/ computational approach illustrated above. That is, they essentially start from a reductionist approach and ask what computations/information processing functions may be performed by these oscillations. *These models have given enormously useful insights into how networks stripped down to basics* perform canonical computations and exchange information (Fries, 2015a; Fries et al., 2007; Jadi and Sejnowski, 2014; Ray et al., 2013; Tiesinga and Sejnowski, 2010; Womelsdorf et al., 2014). In addition, these networks have offered insight into specific computations that may underlie specific parts of higher level aspects of perception, attention and other forms of cognition (Buia and Tiesinga, 2006; Fries, 2009, 2015a; Tiesinga et al., 2005).

3.2.2 Reduced models studying fundamental properties of network oscillations

These models employ some reduction and simplification of model units and network structure in order to address specific questions on the mechanisms by which gamma is generated and maintained. These models of reduced scale and detail have no spatial structure and typically use point neurons to ask basic question on the generation of gamma, and the mechanisms that ensure stability or cause destabilization.

One line of research deals with the network and input conditions under which gamma oscillations arise, what governs their stability, and what destabilises them (Börgers and Kopell, 2005; Brunel and Wang, 2003; Traub et al., 1996; Whittington et al., 1995) (for reviews see (Buzsáki and Wang, 2012; Whittington et al., 2000)). Another line of research focuses on the network parameters that govern key oscillation observables such as their frequency and power (Bartos et al., 2007; Brunel and Wang, 2003; Jia et al., 2013; Mazzoni et al., 2011; Traub et al., 1996; Whittington et al., 1995). Yet another, related line of research deals with identifying the network mechanisms that govern the interaction between extraneous drive, local inhibitory (I) and excitatory (E) populations during the emergence of gamma oscillations. A specific question that has been addressed in this respect is whether the mechanism generating gamma oscillations is an ING (Interneuron Gamma), a strong PING (Pyramidal Interneuron Gamma) or a weak (sparse)- PING (Bartos et al., 2007; Buzsáki and Wang, 2012; Tiesinga and Sejnowski, 2009; Wang, 2010; Whittington et al., 2000).

A major advantage of such relatively reduced models is that, because of a somewhat reduced scale and relative structural simplicity, they are amenable to detailed analysis of simulation outputs and in some cases even to analytical treatment of the model equations(e.g. (Börgers and Kopell, 2005)), which permits the precise identification of mechanistic causes for the observed phenomena. Some of these models have computed oscillation observables in the model (such as gamma frequency and power), which is especially valuable when certain experimental manipulations can be used to approximate changes in specific network parameters (Jia et al., 2013; Mazzoni et al., 2011; Roberts et al., 2013). However, the validity of this comparison depends crucially on the empirical validity of the generative model of the LFP.

Even in reduced models of gamma, many crucial parameters such as connection probabilities, synaptic efficacies and most notably the nature of the input (in terms of its specificity to excitatory and inhibitory populations and its spatial and temporal correlation structure) have limited empirical grounding. These crucial parameters, the choice of which can strongly influence model behaviour and even change the underlying oscillation mechanism (e.g. from ING to weak or strong PING), are often set by convention or are a result of setting some initial values based on literature and 'tuning' parameters such that gamma oscillations are generated. This problem arises mostly because of the lack of sufficient and relevant details in the literature: the experiments to estimate such parameters are very complex and effortful and sometimes even impossible with current techniques. Therefore, one of the key challenges of the Human Brain Project is to find ways to estimate crucial modelling parameters in an empirically validated manner. A further limitation is that the models outlined in this section are typically spatially undifferentiated. In these models, the neuronal connectivities are typically probabilistic (i.e., spatially random), and hence no laminar or columnar structure is modelled. These may be for example models describing the average firing rates in two populations of neurons (inhibitory, excitatory), so called *mean-field (firing-rate or population)* models e.g. (Jia et al., 2013), or models of coupled excitatory and inhibitory neurons (Börgers and Kopell, 2005).

In addition, many studies modelling individual units have employed so-called, *point model-neurons.* Point model neurons have no spatial differentiation into compartments such as the apical dendrites, soma, or distal dendrites; all currents in such model neurons enter and exit the neuron from a single point. For a comparison of the behaviour of different types of point model neurons, see Izhikevitch (Izhikevich, 2004). However, compartmental differentiation is important when attempting to compare a network output observable (e.g., the simulated LFP) with an experimentally observable signal (e.g., the laminar LFP). This is because neuronal cell compartments, especially in large pyramidal neurons, tend to stretch across cortical layers and receive inputs from different origins and with different timings. The resulting transmembrane currents and the associated impressed and return currents typically generate the LFP (Einevoll et al., 2013) and macroscopic signals such as ECoG, MEG and EEG (Hämäläinen et al., 1993; Nunez and Srinivasan, 2006a). Some authors have used LFP proxies in models composed of point neurons (e.g. (Mazzoni et al., 2011; Roberts et al., 2013)). The validity of such approximations is debated (Barbieri et al., 2013; Einevoll et al., 2013). In a highly relevant study on this issue

(Mazzoni et al., 2015), a quantitative comparison was made between various point integrate and fire (IF) neuron LFP-proxies and ground truth biophysical model data (based on multi-compartmental neurons and an appropriate forward model), to evaluate which LFP proxies perform relatively well compared to ground truth data.

The reduced models discussed in this section and the comparison of their output with models composed of multi-compartmental neurons are deepening our insight into the generative mechanisms of gamma. In general, despite the lack of spatial structure, the use of simplified model neurons, and the associated lack of a detailed forward model to generate signals comparable to empirically observed signals, these reduced models given a suitable LFP proxy can benefit from comparison with empirical data. Hence, subjecting these models to empirical validation can only further increase their value for understanding neural processing in (visual) cortex.

4 A perspective on the empirical validation of models

In general, large-scale models, strictly speaking, are difficult to empirically validate. In these models, the number of parameters and combinations of parameter settings that could potentially yield similar functional behaviour is extremely large. However, even in very simplified generative models, that consist of relatively small numbers of point neurons and have no laminar or columnar structure (such as a number of models focused on understanding fundamental generative gamma mechanisms and gamma oscillation properties described previously), the very same problem still holds. There is often not much empirical guidance to set synaptic conductances, levels of adaptation, connectivity strength among inhibitory, among excitatory, and between excitatory and inhibitory neurons, etc. Unfortunately, the question of which of the parameter configurations is most likely to yield model data that may underlie the empirical observations is difficult to answer. This is because many of the network parameters that are crucial for its behaviour such as the parameters specifying the input and the parameters specifying the effective neuronal connectivity and many other aspects are unobserved, and often set by convention. At the same time, even relatively small variations of these parameters, strictly within biologically plausible limits, may lead to qualitatively different behaviour.

However, simplified models of reduced scale and/or with simplified units also have important limitations. A recent study showed that when comparing, otherwise similar networks with either more computationally-simple, current-based synapses, with more realistic, conductance-based synapses, the second order statistics of neural population interactions in the network (such as spike train correlation) and their input-related modulation were different (Cavallari et al., 2014). Furthermore, activity of networks with the more realistic conductance-based synapses showed stronger synchronization in the gamma band, the spectral features of which also carried more information about the input. Thus, even after applying simplifications such as removing spatial structure in the network, and using point neurons rather than multi-compartmental biophysical neurons, the specific choice on the type of synapse can be crucial for the performance of the model. In this case, the crucial choice pertains to employing more realism in the form of conductance-based synapses and perhaps also conductance-based Hodgkin Huxley (HH) rather than more abstract integrate-and-fire (IF) model neurons.

Having settled on what is the minimal realism necessary for addressing the question at hand and having built such reduced model, the next question to be addressed is which of the configurations of largely unknown parameter values is most likely to underlie the data. To solve this problem, we suggest a stepwise process of empirical validation (Zachariou et al., 2015). In this empirically constrained modelling approach, relatively simple (e.g. HH type) point neurons are used at first in small spatially-undifferentiated networks in which however the parameter space is maximally validated by empirical data at the correct scale. Such an empirically validated model aims at the characterization and extraction of robust and valid descriptors of network behaviour for a specific experimental setting. This approach is to be maintained when upscaling the network in terms of size and complexity towards from meso- to macro-scales. In a full scale model, it is then required that all constraints from micro-, meso- to macro-scale are satisfied simultaneously. In addition, the simultaneous availability of data at multiple scales is important to help constrain their interactions in the model. For example, the precise link between micro-scale (as quantified from sorted single unit spike trains) and meso-scale (population activity as measured by LFPs) will determine the underlying generative models, which include ING, PING, or weak-PING (Wang, 2010; Whittington et al., 2000). Hence, it is crucial that empirical constraints are quantified and implemented at both micro- and meso-scopic scales, and thus should address both single unit and LFP behaviour. To constrain models at the macro-scale, additional empirical constraints from ECoG and MEG measures are necessary. Irrespective of the scale of modelling, empirical validation is carried out by running the model at all different settings of unknown parameters, and to choose the configuration that best matches the empirical data corresponding to the scale of modelling. The model, whose behaviour shows the best match to empirical constraints, is an *empirically-validated model*. Such a model network with sufficient (but perhaps not extensive) realism can then be analysed thoroughly to gain insights into the underlying mechanisms for the phenomena of interest.

The obvious advantage of this compromise-approach is that the empirical validation may be a more tractable problem than in generic large scale and or complex models, because the parameter space that requires estimation is much smaller. If (quasi)quantitative empirical validation is aimed for, then this reduction in parameter space in more limited models is of paramount importance because it is difficult to estimate too many parameters from the finite and noisy datasets that are typically recorded experimentally. The disadvantages of these models are related to the insufficient realism in

term of scale/neuronal density, which in turn may lead to not attaining the full network dynamics as for the results of (van Albada et al., 2015) and thus not attaining the true correlational structure - leading to a poor model LFP estimation (Hagen et al., 2015). If point neurons are used, limitations with respect to attaining a realistic LFP, readily comparable with experimental data apply (Einevoll et al., 2013; Mazzoni et al., 2015). Nevertheless, there may be ways to estimate LFP from point neurons (Mazzoni et al., 2015). Furthermore, introducing a hybrid approach in which point neurons are mixed with multi-compartmental neurons in differentiated models (Hagen et al., 2015) can provide a workable approach to find an acceptable compromise between empirical validation and sufficient generative realism.

Gamma oscillations are particularly suited to test our approach of stepwise, empirically validated modelling. The main reason for the suitability of gamma oscillations to this end are fivefold. First, gamma oscillations are important in terms of the *psychophysiological approach* (section 5), demonstrating statistically-reliable and experimentally-robust associations with mental functions especially in sensation and perception (for a review see (Fries, 2009)). Second, gamma oscillations have been implicated in theories addressing aims of the functional/computational approach (Section 5). Third, the generative network mechanisms of gamma oscillations have been studied widely in-vitro and in-vivo, and have been the target of extensive mathematical/computational modelling (for reviews see (Bartos et al., 2007; Buzsáki and Wang, 2012; Tiesinga and Sejnowski, 2009)). Fourth, gamma oscillations have been studied in detail in the early visual cortex of the human and non-human primate. This is important because the functional anatomy/physiology of the visual system (and early visual cortex especially) is well understood, and because the sources of gamma activity both macroscopically (MEG and to a lesser extent EEG) and more mesoscopically (ECoG , LFP and laminar LFP, CSD) have been elucidated to a great extent (see (Hadjipapas et al., 2015). Fifth, key functional (for instance stimulus-related) behaviour of gamma has been well-documented at the different spatial scales and species (human vs. nonhuman primate), as will be further shown in the next section.

5 Gamma: A case in point for empirical validation of models

Neurons communicate predominantly with each other through spiking, but excitability (and hence spiking probability) tends to vary cyclically over time. This implies that the presence or absence of a favourable temporal relationship among the excitable periods in different populations can make or break their communication (Fries, 2005a, 2009, 2015a; Fries et al., 2007). These cyclical or oscillatory changes occur not only in the gamma frequency range (25-80Hz), but also in several other frequency ranges referred to as beta (14-25Hz), alpha (8-13Hz), theta (4-7Hz), and delta (1-4H) (Clayton et al., 2016). Here, we will focus on the role of gamma oscillations in neuronal communication within V1, and early visual cortex. Despite what seems to be robust evidence for a role of gamma in stimulus processing and neuronal communication, there has been a persistent counter-view in which gamma is seen as a predominantly epiphenomenal feature reflecting the architecture of cortex, but without functional implications. We suggest that this debate has persisted due to the lack of predictive generative gamma models (i.e. models that can predict in which conditions gamma does or does not occur).

5.1. The gamma debate

Until recently it was thought that gamma frequency was highly stable over time and across brain areas in a given individual (Hoogenboom et al., 2006; Muthukumaraswamy et al., 2010), which would ensure efficient communication among remote neuronal populations. However, especially in recent years, evidence has accumulated that gamma frequency can vary considerably. From a theoretical/ modelling perspective, the idea that gamma frequencies are by default matched across different brain areas is not that straightforward, as gamma is often assumed to depend on local network properties (Buia and Tiesinga, 2006; Fries, 2005a), which may differ among areas. In addition, strong dependencies of gamma band frequencies on visual stimulus parameters have been demonstrated (Feng et al., 2010; Gieselmann and Thiele, 2008; Ray and Maunsell, 2010; Swettenham et al., 2009), which may interact with differences in local mechanisms generating gamma in different areas. Strikingly, Roberts et al (Roberts et al., 2013) demonstrated a >20Hz shift in gamma frequency with stimulus contrast in V1 (Figure 9, also see (Jia et al., 2013; Ray and Maunsell, 2011)

This seemingly simple finding has had a large impact on the theoretical understanding of gamma in early visual cortex: In the domain of gamma oscillations, two important theories have been proposed that address the contribution of gamma to neuronal communication. The first one has become known as the binding theory. This theory proposes that different neurons in the brain that are encoding different spatial loci as well as different features of an object are bound together by gamma to form a coherent percept (Engel et al., 1999; Grossberg, 1976; Von Der Malsburg, 1994; Milner, 1974; Singer, 1995). According to the maximal version of this theory, all relevant neurons in the brain encoding an object would be bound together by having an appropriate match (phase relationship) among excitable periods occurring in the gamma frequency range. A similar mechanism may be envisaged to form neural assemblies representing retrieved memories, actions while they are being executed, or feelings while they are being experienced. In all of these cases, large numbers of neurons are proposed to form functional networks that interact within and across different cortical areas as well as subcortical structures. In the context of the present review, we will use the term binding only to refer to interactions within V1, which we more generally will refer to as within-area communication (WaC).

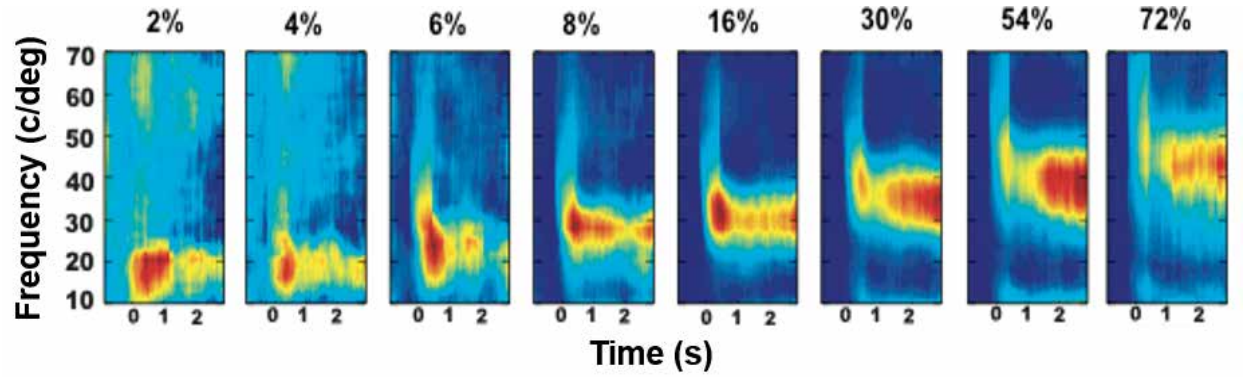


Figure 9: Average data from a single monkey showing the increase in gamma frequency as a function of grating contrast in V1.

Each panel represents gamma power (relative to baseline) as a function of time (stimulus onset at zero). Data replotted from Roberts et al. (Roberts et al., 2013)

A second highly influential theoretical framework in the domain of neuronal communication in the gamma range is the communication through coherence (CTC) theory by Fries (Fries, 2005a, 2015a; Fries et al., 2007). In CTC, important principles have been proposed about the conditions under which communication between areas can occur among different cortical areas. The main idea, that an appropriate match between windows of excitability is required, is shared between binding and CTC theories. When limiting oneself to early visual areas, the difference between the two theories might be related more to the types of phenomena they have focused on, than on the basic principles of neural communication. Binding can be seen as more related to the way in which local processes within an area contribute to the formation of a Gestalt from local elements (Engel et al., 1991; Gray et al., 1989), and figure-ground segregation (Lamme, 1995; Peterson and Salvagio, 2009; Poort et al., 2012; Self et al., 2013), whereas CTC has been used primarily to understand the selective routing of information between areas at different hierarchical levels, especially in studies of selective attention (Bosman et al., 2012; Fries, 2005a; Salinas and Sejnowski, 2001; Tiesinga et al., 2002; Wildie and Shanahan, 2011). Nevertheless, the two theoretical frameworks are highly related, as illustrated by studies showing that binding is facilitated by attention (Ashby et al., 1996; Shafritz et al., 2002; Treisman, 1998, 2004; Treisman and Gelade, 1980). By analogy we use the term within-area communication by synchronization (wCS) rather than binding and the term between-area communication by synchronization (bCS) rather than CTC. The reason for the latter is that the term 'coherence' refers to a specific method of computing synchronization that is not applicable when the empirical data do not satisfy specific constraints (Lowet et al., 2016). Note furthermore that in line with considerations by Fries (Fries, 2015a), we will define the synchronization between neural populations as referring to the process of optimizing communication by means of arranging windows of excitation in two populations at a time (phase) difference appropriate to compensate for transmission delays. In the present review, we want to discuss challenges of communication by synchronization, and how these challenges have led to new empirical, methodological, and computational research that can inform modelling of early visual cortex.

The finding that gamma frequency depends on stimulus properties implies the possibility for frequency mismatch resulting in an input-dependency of the efficiency of information transfer between populations. This would render synchronization as a mechanism for neural communication implausible (Jia et al., 2013). A highly relevant finding to evaluate the contribution of gamma to wCS came from Ray and Maunsell (Ray and Maunsell, 2010). They recorded from pairs of V1 neurons responding to different parts of a Gabor stimulus, responded with differing peak-power gamma frequencies to different contrast regions. Importantly, gamma synchronization was reduced as a function of the contrast difference between different stimulus location. This finding was interpreted as evidence against binding-by-synchrony, because they showed that there was no evidence for matching gamma spectra within the V1 representation of the object. Moreover, several authors (Burns et al., 2011; Roberts et al., 2013; Xing et al., 2012), have reported large seemingly random gamma power and frequency (~15Hz) modulations in V1, in the absence of changes in stimulation. These data suggest that for neural communication by synchrony to work, there must be mechanisms not only to determine appropriate phase differences, but also to match frequencies sufficiently to enable sufficiency stable phase relationships.

5.2 A computational view on neural communication by synchrony in V1

The discussion in the literature about whether gamma synchronization can play a role in neural communication may so far have been largely characterized by a lack of computational insight into the mechanisms that regulate frequency, phase and power differences among communicating neural populations. To address this, it is useful to first consider the neurophysiological mechanisms by which gamma is generated. Gamma is thought to be generated by interactions between 'regular-spiking' pyramidal cells and 'fast-spiking' inhibitory cells. Among inhibitory cells, basket cells are thought to play a prominent role, because of their

powerful perisomatic targeting of inhibition onto pyramidal cells. Among the possibilities for generating gamma, two have been prominently discussed (Buzsáki, 2006; Tiesinga and Sejnowski, 2009). In the first scenario, basket cells, which communicate with each other synaptically and through gap junctions form a network in which randomly occurring patterning of spiking generates a gamma rhythm. With proper connectivity and assuming fast-acting and decaying inhibition (by $GABA_A$) frequencies in the gamma network can be expected in the gamma range, which in turn may modulate the excitability of pyramidal cells at that same rhythm. This model is referred to as the InterNeuron Gamma (ING) model. In an alternative scenario, the rhythm is generated by excitatory drive to the pyramidal cells, which excites local basket inhibitory neurons that in turn given inhibitory feedback to the pyramidal cells. Here, the gamma rhythm will depend on the interplay between the strength of the excitatory drive and the decay time for inhibition. This mechanism, referred to as the Pyramidal InterNeuron Gamma (PING) model, is interesting in the context of findings that gamma is stimulus-dependent, and indeed computational PING models exhibit stimulus-dependency (Hadjipapas et al., 2015; Roberts et al., 2013), see Figure 10). In addition to the ING/PING distinction, it is important to consider how local the network is that generates the gamma frequencies. Assuming a strongly connected inhibitory network in the ING model, large networks may converge upon a single narrow gamma spectrum, entraining pyramidal cells at different phases (Buzsáki and Wang, 2012; Fries et al., 2007; Lisman and Jensen, 2013). In this case, the inhibitory network acts as a single clock to which activity from other cells is referenced. Alternatively, the connectivity among inhibitory (and excitatory) neurons may not be set up to permit the long-range spreading of oscillatory activity, which could give rise to multiple clocks (Buzsáki and Wang, 2012; Fries et al., 2007; Lisman and Jensen, 2013).

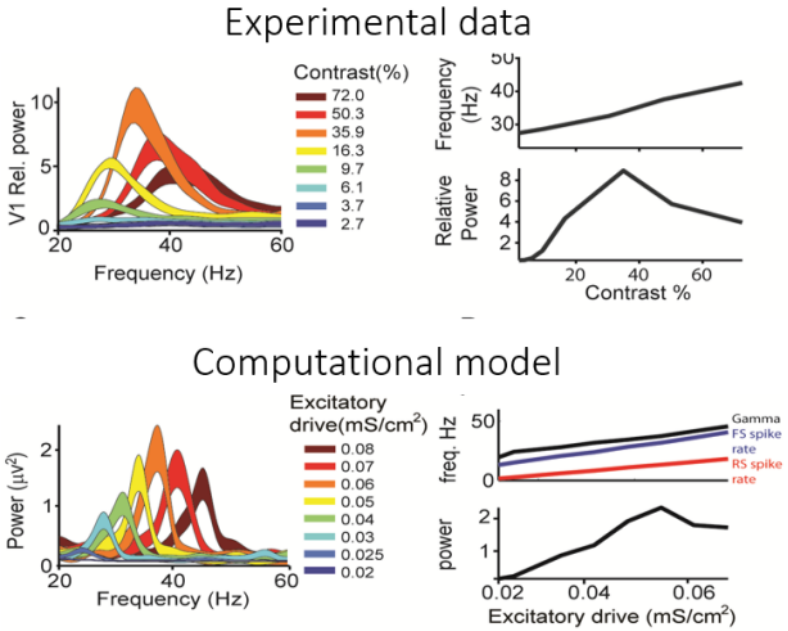


Figure 10

Top: Empirically obtained gamma spectra as a function of grating contrast in V1 of a single monkey, with on the right hand side estimated peak-power frequency as well as relative power as a function of contrast. *Bottom: Analogous data obtained from PING modelling. Spectral responses are shown on the left and frequency of gamma, fast-spiking (FS) inhibitory neurons (blue), regular spiking (RS) excitatory neurons (red), and power are plotted as a function of excitatory drive (thought to be a proxy of contrast). Experimental data replotted from Roberts et al. (2013).*

Here, we will consider PING networks in which the inhibitory and excitatory cells receive local excitatory drive (stimulation), and whose gamma oscillations are generated locally. In this case, a PING network can be considered as an oscillator, and the question can be asked how these oscillators interact. Anatomically, PING networks may be compared to local populations in superficial layers in V1, where pyramidal cells are known to show lateral connections extending over up to 5mm (Yoshioka et al., 1996). Hence, the interaction between PING networks can be approximated by an interaction between oscillators. Mathematically, the synchronization of interacting limit-cycle oscillators is well understood (Ermentrout and Kopell 1984) (Ermentrout and Kleinfeld, 2001; Hoppensteadt and Izhikevich, 1996, 1998). Here, we will focus on the theory of weakly coupled oscillators (TWCO, for review see (Pikovsky et al., 2002) which has been applied in a broad array of scientific domains including neuroscience (Bendels and Leibold, 2007; Breakspear et al., 2010; Ermentrout and Kleinfeld, 2001; Galán et al., 2005; Hoppensteadt and Izhikevich, 1996, 1998). In TWCO the phase of an oscillator is defined by an intrinsic (natural) frequency, and the interaction with

other oscillators is characterized by the phase response curve (PRC, for review see (Schwemmer and Lewis, 2012)). The PRC defines how the phase of an oscillator is modified by its interaction with other oscillators. The amount of phase-locking (the strength of synchronization) between oscillators depends on the interaction (coupling) strength between oscillators and their intrinsic frequency difference (also referred to as detuning). The resulting interplay of detuning and coupling is expected to define a triangular region of synchronization in a coupling-versus-detuning space (Coombes and Bressloff, 1999; Pikovsky et al., 2002; Tiesinga and Sejnowski, 2010). This triangular region is often referred to as the 1:1 Arnold tongue (Figure 11). In TWCO, the coupling strength is 'weak', meaning that the interaction among oscillators mainly affects phases and frequencies, but not their oscillation amplitudes.

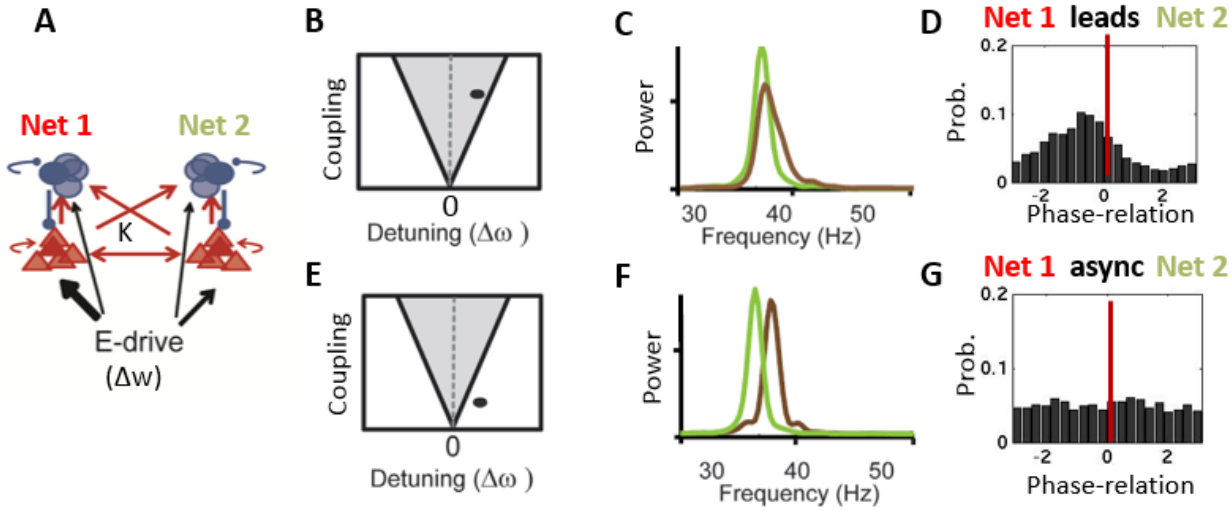


Figure 11: Understanding the interaction and synchronization between PING networks in the framework of TWCO.

A) Two PING networks are constructed and coupled (excitatory cells and connections in red; inhibitory cells and connections in blue), after which Net 1 is given stronger excitatory drive (fatter black arrow) than Net 2 (thinner black arrow). B) 2D-space of coupling strength versus detuning. Black dot falls within the Arnold tongue, indicating that there is sufficient coupling to achieve a given measure of synchronization for the imposed detuning (intrinsic frequency difference). C) The two nets achieve a similar narrow-band gamma spectrum despite differential excitatory drive. D) The higher intrinsic frequency of Net 1 than for Net 2 after synchronization is translated into a phase difference, with Net 1 leading Net 2 (with intrinsic frequencies referring to the frequencies obtained if Nets 1 and 2 were uncoupled). E) A reduction of coupling strength for the same level of detuning. F) Lack of synchronization when coupling is too weak given the level of detuning. Grey region in B and E represents the expected regions within which synchronization would occur should all coupling vs detuning combinations be tested. G) Lack of synchronization leads to a lack of a fixed phase relationship between oscillators and to phase precession. Redrawn from Lowet et al., 2015.

Lowet et al. (Lowet et al., 2015a) completed a study in which for a pair of mutually interacting PING networks, synchronization properties were tested for a large range of combinations of coupling and detuning. Compared to Tiesinga and Sejnowski (Tiesinga and Sejnowski, 2010), who first applied TWCO in a realistic gamma network for visual cortex, Lowet et al. (Lowet et al., 2015a) went several steps further. They studied a spatially-continuous PING gamma network in which local connectivity decayed as a function of spatial distance, and in which different network locations received a different strength of input drive. As a result, gamma synchronization could be kept local within the network in accordance with findings from (Ray and Maunsell, 2010). In addition, Lowet et al. (Lowet et al., 2015a) used input drive ranges in the model that induced a range of gamma oscillation frequencies that matched empirical observations from Roberts et al. (Roberts et al., 2013). An investigation of a large range of coupling and detuning conditions led to observations of phase-locking, phase-relations and frequencies among PING models in line with TWCO and the Arnold tongue. Figure 12A shows the predicted triangular region of phase-locking (synchronization). Furthermore, phase and frequency coding of input was largely complementary, again in line with TWCO (Figure 12B, C). Large differences in input strength, being largely outside the Arnold tongue, are encoded as frequency differences. Finer differences in frequency within the triangular region of synchronization are translated into phase differences. In line with the distinction between encoding of coarse and fine input differences by respectively frequency and phase, combining the two codes yielded the best reconstruction of stimulus input. The behaviour shown in Figure 11 and Figure 12 is robust against a range of variations in the PING model (e.g., more or less sparse firing in the excitatory neurons), is not dependent on the type of model neuron (e.g.,HH vs Izhikevich), and the behaviours among coupled PING networks can also be seen among coupled oscillators (see Lowet et al. (Lowet et al., 2015a)).

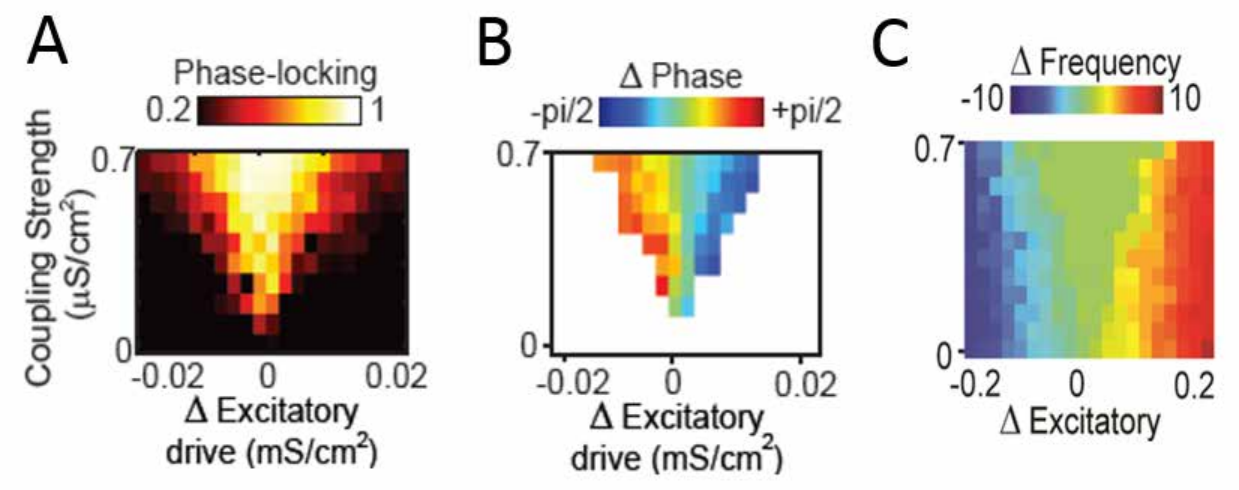


Figure 12: Properties of the Arnold tongue

(A) Region of synchronization in the coupling-by-detuning space (detuning shown here as a difference in excitatory drive). Synchronization is quantified by a phase-locking index. B). within the region of good phase locking, the locking occurs at different phase differences, such that the network with the higher intrinsic frequency will lead in phase. C) Illustration of the common frequency upon which oscillators with different intrinsic frequencies will emerge for the proper combinations of detuning and coupling. Based on Lowet et al. (2015).

To demonstrate that TWCO is a useful approach to understand interactions in topographic visual networks, Lowet et al. (2015) constructed a 100 by 100 array of phase oscillators, in which oscillators were connected with their neighbours with a strength and probability decaying exponentially with distance, as has been documented in early visual cortex (Figure 13). In line with conductional delays for horizontal connections on the order of ~0.3 mm/ms in V1 (Angelucci and Bullier, 2003; Boucsein et al., 2011), Lowet et al. (Lowet et al., 2015a) used a time-delay term in the lattice model that increased as a linear function of spatial distance (slope of 0.4 ms/pixel, offset 1ms). Conduction delays affect phase-relations as well as phase-locking, and limit the extent of spread of synchronization in a topographic neural network. This network was then exposed to the contrasts extracted from natural images, by assigning an intrinsic frequency to each oscillator in line with its corresponding image contrast. Upon stimulation, the network showed interactions among oscillators, with the emergence of synchronization fields, which appeared to have useful properties for grouping and figure-ground segregation. In particular, although the emergent synchronization fields often did not fill complete surfaces, they showed the useful property of rarely crossing object boundaries. This can be understood from the perspective that contrast variance is greater at object boundaries than within the surfaces of these objects. This facilitates the emergence of synchronization fields within object surfaces, whereas the often large contrast discontinuity at borders will have the tendency to break topographic synchronization regions.

The potential of oscillating neural networks for meaningful segmentation of input patterns has been described before (Chen and Wang, 2002; Kuntimad and Ranganath, 1999; Kuzmina et al., 2004). However, in some studies, the clustering is based on a phase-code only (Eckhorn, 1999; Kuntimad and Ranganath, 1999; Wang and Terman, 1995, 1997), whereas in others it is mainly based on de-/synchronization (König and Schillen, 1991). In Lowet et al. (Lowet et al., 2015a), local input differences in a network are translated in a combined frequency and phase code. Importantly, the nature of gamma in our modelling approach is local. This makes the modelling approach fundamentally different from other model architectures characterized by global synchrony, like the LEGION model (local excitatory global inhibitory oscillator network, (Wang and Terman, 1995)) or the PCNN (pulse-coupled neural network, (Kuntimad and Ranganath, 1999)). In these models, clustering was based on phase alone and the network had a single main frequency. LEGION and PCNN perform well in image segmentation tasks, but are not biologically plausible. The idea that there is information in the frequency and phase relations of oscillatory responses that is relevant for reconstructing input, does not prevent other types of encoding to be relevant as well. There are various types of information in the spiking of neurons that can be exploited while making abstraction of oscillatory properties. Likely, the brain will combine different types of codes depending on which code is more efficient given specific stimuli and behavioural requirements. Indeed, each type of encoding has its own advantages and disadvantages. For example, spiking rate ('rate-coding hypothesis') may sometimes relate closely with changes in stimulus parameters, but it requires sufficient integration over time, and its range of encoding can be limited by neuronal saturation. On the other hand, precise spike timing contains significant information about the stimulus (Masquelier et al., 2009; Rieke et al., 1997; VanRullen et al., 2005), however only short integration time windows are required to extract that information. Spike timing can also be considered in neuronal populations, an idea that has led to the relative spike-timing hypothesis (König and Schillen, 1991; Sakurai,

1996; Singer, 1999; Tsukada et al., 1996). Information is represented here in the exact spiking pattern of a number of neurons.

An important consideration is that synchronization usually does not lead to complete phase locking. Instead, noise and state-related factors will usually lead to partial (intermittent) synchronization regimes. Even without these factors, TWCO predicts that in almost all conditions (except when detuning is zero to begin with) partial synchronization is the rule rather than the exception. This is supported by reports of significant strengths of synchronization among neuronal populations with non-matching narrow-band gamma spectra (Bosman et al., 2012; Gregoriou et al., 2009; Ray and Maunsell, 2010). This suggests that these populations alternate periods of phase locking (with a matched frequency) with periods of phase precession. The non-stationarity of the resulting oscillations precludes the use of the widely used coherence index (Oostenveld et al., 2011) if the aim of analysis is to assess phase relations among neural populations. Instead other approaches to estimate phase relations among neural populations have been proposed, in which methods of frequency band selection are followed by the Hilbert transform for estimating instantaneous phase (Lowet et al., 2016). The latter approach may be preferable for the analysis of empirical data as well as modelling output if the goal is to obtain estimates of phase and phase-relations among neural populations.

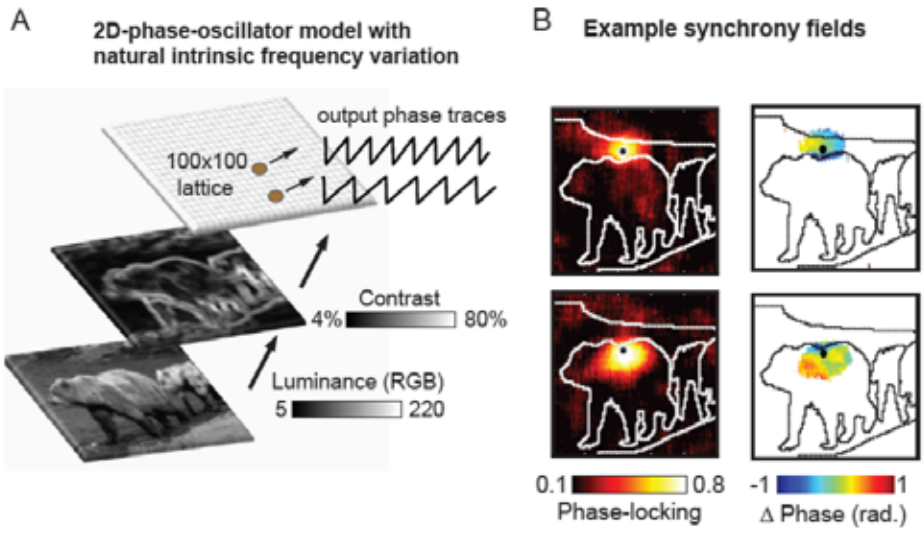


Figure 13: Phase-oscillator model with natural image input.

A) Image processing. Each natural image was reduced to 100x100 pixels and transformed from a luminance image into a contrast image. Data from Roberts et al. (2013) was used to transform local contrast values into intrinsic gamma-frequencies per phase oscillator in the 100x100 lattice network. B) Two examples of emerging synchrony fields (color) defined by comparing a reference oscillator (black dot) to all other oscillators. In the resulting emerging synchronization field, color coding reflects either phase locking (left column) or phase difference (right column). When the reference (black dot) was located outside of the main object (top row), the resulting synchronization field did not extend into the object. Likewise, when the reference (black dot) was located inside of the main object (bottom row), the resulting synchronization field remained inside the object surface.

The V1 modelling of our group reviewed here represents an important step forward in moving from a descriptive analysis of synchronization towards a predictive analysis based on theoretical principles. The demonstration that TWCO guides oscillatory interactions in the gamma range in V1 is a definitive step forward in developing more founded computational model of V1 gamma synchronization. Carefully studying these limited models can yield useful information into the principles behind neural architecture. However, the value of these insights is currently limited by the biological plausibility of the model. In order to increase our confidence that our ideas are indeed relevant, further work is required to make the PING models we have used more realistic (by including more realistic structure and connectivity), and to carefully test predictions of TWCO in these networks against empirical data. In ongoing work (Lowet et al., 2012), we are analysing relevant empirical data from V1 obtained with multiple laminar probes, at different distances from each other, and each stimulated with different local contrast. Furthermore, the models need to be expanded to include interactions among layers, and between different areas. Careful work is also required to calibrate models to take into account individual differences which reflect individual genomic differences (van Pelt et al., 2012). Moreover, it is possible that PING models which accurately reflect non-primate cortex would have to be adjusted to account for differences between species.

5.3. V1 – V2 gamma synchronization: Empirical and modelling data

To test whether gamma synchronization can be a mechanism for information transmission from V1 to V2 Roberts et al (Roberts et al., 2013) recorded simultaneously in V1 and V2 with depth probes, with recording contacts spanning the cortex from superficial

to deep layers in both areas. Figure 14A (left panel) shows there is a pattern of synchronization that preferentially links sites in superficial layers of V1 with layer IV in V2. The distribution of the observed pattern of synchronization was in line with the anatomy of feedforward anatomical connectivity (details in (Roberts et al., 2013)). Importantly, the pattern of functional connectivity from V1 to V2 during visual stimulation was constant with contrast. Figure 14A shows data from two monkeys that showed similar patterns of connectivity for a higher and a lower contrast. Hence, although different contrasts induce different gamma frequencies, synchronization between V1 and V2 remained possible. Furthermore, Granger causality analysis showed that the observed pattern of synchronization was associated predominantly with a feedforward direction of information transmission. These data fit well with the view that gamma is involved in feedforward information transmission, between LGN and cortex (Bastos et al., 2014), and within visual cortex (Bastos et al., 2015a; Bosman et al., 2012; Fries, 2015a; Zandvakili and Kohn, 2016).

It is remarkable that despite strong variability in gamma power and frequency visible in individual trials (Figure 14B, bottom panel), V1 and V2 were able to respond to stimuli with a robust level of synchronization. In line with observations from (Burns et al., 2011; Xing et al., 2012), Roberts et al. (Roberts et al., 2013) found a large variation of ~15Hz in instantaneous frequency (for constant stimulation conditions). However, there was also a strong correlation between instantaneous frequencies in V1 and V2 (Figure 14C), which led the authors to suggest the existence of a mechanism for realizing sufficient a frequency match between the two areas. The frequency matching permitted V1 and V2 to show synchronization in spectra that shifted with contrast in a manner resembling the shifts of gamma power spectra in V1 and V2 individually (Figure 14D). A computational model in which two PING models were coupled to mimic the feedforward connectivity between superficial V1 and layer 4 in V2 showed that the V2 model network became entrained by the frequency of V1, so that V2 gamma was similar for V1 gamma for different level of excitatory drive to V1 (Figure 14E). Roberts et al. (Roberts et al., 2013) emphasized the need for achieving a sufficient frequency match during the entrainment, and this view is in line with the application of insights from TWCO to gamma synchronization discussed in the previous section.

Since individual trials show gamma bursts rather than sustained gamma, the entrainment must occur very rapidly, at the time scale of these bursts. In Figure 14B, the black lines correspond to the peaks of a theta rhythm that was present concurrently in the single trial oscillatory data. Lowet et al. (Lowet et al., 2015b) followed up on these observations and found that the theta rhythm in V1 and V2 was generated by microsaccades occurring while the monkeys were fixating the fixation spot. Interestingly, they found that gamma synchronization between V1 and V2 occurred at the rhythm of microsaccades. The idea that long-range neural communication depends on rhythms initiated through actions by the sensory organ is related to the theoretical concept of active sensing (Schroeder et al., 2010; Tomassini et al., 2015). The influence of microsaccades is not limited to the visual system; it has been reported for example that saccades influence the hippocampal theta (Hoffman et al., 2013), which in turn structures gamma into bursts nested in the theta rhythm (Jutras and Buffalo, 2010). The patterning of feedforward information transmission in gamma bursts by slower rhythms is interesting, because it puts constraints on the way in which feedback by slower rhythms such as alpha or beta (Bastos et al., 2015b; Buffalo et al., 2011; Fries, 2015a; van Kerkoerle et al., 2014; Michalareas et al., 2016; Zandvakili and Kohn, 2016) influence feedforward transmission via gamma. Presumably, the entire interaction between feedforward and feedback influences must happen within the time window of a gamma burst. A systematic analysis of gamma bursts and their relation with microsaccades by (Lowet et al., 2015b) showed a strong tendency for gamma frequency to start higher at the beginning of a burst and then to decline. This decline is in line with the view that (bottom-up) drive after each microsaccade declines over time, making it likely that the effect of feedback would have a stronger relative impact on the later part of the gamma burst. This is also in line with spiking data (Lamme, 1995; Lamme and Roelfsema, 2000; Reynolds and Desimone, 1999) showing that feedback-influences related to figure-ground segregation or attention maximize their influence after the initial burst of feedforward activity (here from stimulus onset) had subsided.

5.3. Using gamma to validate cross-scale visual cortex models in human and non-human primate

In a recent study of gamma oscillations (Hadjipapas et al., 2015), it was shown that principled cross-species and cross-scale (single-unit, LFP, MEG) comparisons are feasible. This comparison yielded marked similarities across scales and species in the functional behaviour of gamma frequency as a function of stimulus contrast. Thus, the gamma frequency response to contrast (input strength) is robust across scales and species and likely provides a signature for network response to input. At the same time intriguing dissimilarities were observed in gamma power between MEG compared to LFP and single unit spiking data (Figure 8).

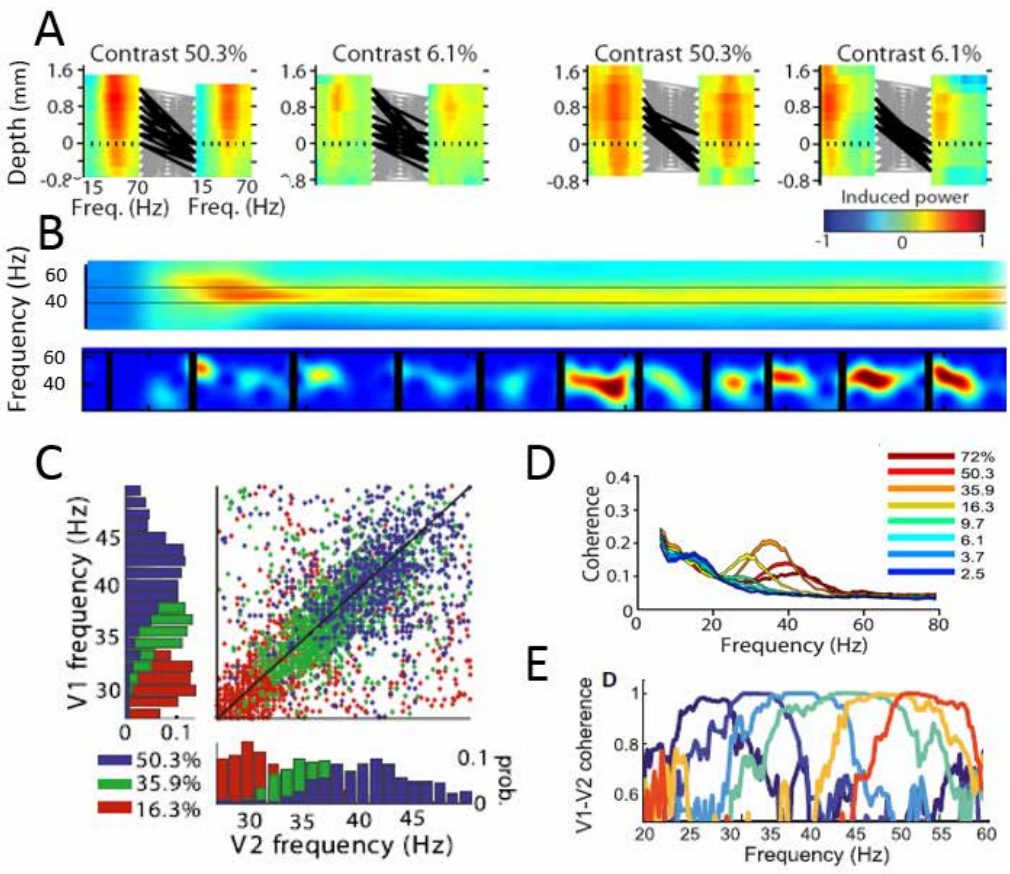


Figure 14: Robust V1-V2 synchronization despite the bursty nature of gamma.

A) *Superficial (V1) to deeper (V2) pattern of synchronization induced by gratings. Each of the 4 subpanels shows within V1 (left and V2 (right) the power of gamma (color-coded) as a function of depth (y-axis) and as a function of gamma frequency (x-axis). The horizontal dashed line corresponds to the estimated top of layer 4. Gamma power is strongest in superficial layers, and the peak power depends on contrast (compare 50% vs 6.1%). The grey lines show all possible pairing between depth probe contact points in V1 and V2 cortex used for estimating synchronization (coherence), and black lines show the top 5% strongest values. The pattern of strongest coherence was not affected by contrast, and was similar for two monkeys (compare the two leftmost panels with the two rightmost panels).* ***B)*** *Gamma is bursty. Top panel shows the average TFR of 50 trials (2secs), revealing the characteristic sustained gamma following stimulus onset reported in many papers. Bottom panel shows that TFR of a single one of these 50 trials. Sustained gamma is an averaging artefact. Black lines show delta peaks, to which gamma bursts are aligned.* ***C)*** *Gamma frequency is variable. Both V1 and V2 show very large variations of instantaneous frequency in the range of 15-20Hz at constant levels of contrast (coded in red, green and blue). This is shown in the marginal distributions of the scatter plot, which reveals the tight trial-by-trial correlation of gamma instantaneous frequency.* ***D)*** *V1-V2 synchronization (estimated by coherence). The magnitude of coherence reflects the gamma spectra, and their shift as a function of changes in contrast (see legend).* ***E)*** *Replication of V1-V2 coherence by linking two PING models. V1-V2 model coherence as a function of excitatory drive to the V1 model closely mimicked empirically observed coherence. Data from* (Roberts et al., 2013).

The understanding of differences in gamma power may give clues on various sources of differences between scales of measurements, for instance in terms of laminar differentiation and associated signal observability biases but also crucially in lateral connectivity/synchronization (Musall et al., 2014). Thus, such dissociations in the functional behaviour of different oscillation observables under the same experimental conditions point the way towards formulating further concrete research questions. One such research question is what generative mechanism *governs gamma oscillation power in the presumably more local and structurally simpler networks generating laminar LFPs, as opposed to the more complex mechanism generating the more global MEG signal.*

We suggest that one way forward in constructing an empirically validated cross-scale model of visual cortex will come from a step-wise approach. A first important step is to create a gamma generating model at the micro and meso-scale that is empirically validated by data at the proper level (LFP and spikes). In further steps, multi-layer and laterally expanded networks can be formed, in which proper within- and across-layer connectivity must be implemented. These models will yield generative forward models of more global signals, which can be empirically validated against ECoG or MEG data. By increasing the scale of the model, new factors will come into play and possibly specific parameters or other model properties will have to be adjusted. Importantly, the effective generators of the signals at different scales in a cross-scale model of visual cortex may not be invariant but may depend on the nature and spatial

distribution of the input, and on brain state (Łęski et al., 2013; Lindén et al., 2011). In addition, the present review has not taken into account the tuning of neurons in early visual cortex to various parameters, and their arrangement into different functional domains. As a concluding thought, by focusing our review on gamma we do not wish to claim that it corresponds to the only or even the most important process in visual cortex. However, we do suggest that building models of increasing scale and complexity, while using gamma as one of the tools for empirical validation can be a productive way forward to generate full-scale models of visual cortex.

6 From V1 to the rest of the brain

In the present review, we have focused on modelling of early visual cortex (mostly V1). The advantage of V1 is that it is probably the best-studied region of neocortex. The amount of known functional-anatomical detail is so large that it has been possible to build structurally realistic models, with the downside that a validation of network parameters with empirical data is difficult due to the large parameter space that would have to be fitted. As an alternative, we have proposed a modelling approach that focuses on the modelling of relevant functional phenomena (e.g., figure-ground segregation, tuning properties, gamma oscillations), and build models that are more abstract, but which give the advantage of a strongly reduced parameter space that permits empirical validation. This difficult balancing act between structural realism and functional validation of models as described here for V1 is representative for the modelling of the whole brain. Empirically validated modelling of sensory areas is also an important aspect of the larger project to model the brain, as sensory areas do not only provide the input to the rest of the brain, but are also strongly involved in the broader brain networks that sustain high-level cognitive operations, including attention (Posner and Gilbert, 1999), working memory (Supèr et al., 2001), imagery (Klein et al., 2004), as well as object categorization and recognition (Cichy et al., 2014). Having a well-validated model of sensory areas will be important to specify the layer-specific and cell-specific connectivities between the sensory areas on the one hand and higher-level sensory, subcortical structures, and association cortices on the other, which are essential for normal brain function.

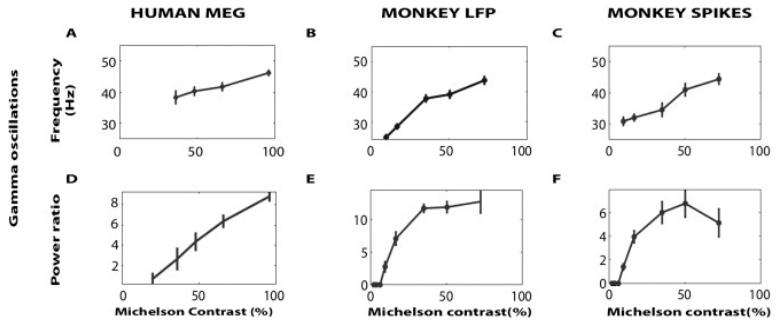


Figure 15: Quantification of gamma spectral effects in humans and monkeys.

A-C) The gamma peak frequency is plotted as a function of contrast for human MEG data (A) monkey LFP data (B) and monkey spiking data (C). D-F) Gamma power plotted as a function of contrast for human MEG data (D), Monkey LFP data (E) and monkey spiking data (F). For more details and additional analysis, see (Hadjipapas et al., 2015)*. The main similarity across scales concerns the gamma peak frequency dependency on contrast. The main difference observed concerns the gamma power dependency on contrast. While power decay and power saturation is observed at high luminance contrasts in LFP and single unit spikes, no saturation is present at high contrasts leading to linear scaling in human MEG gamma power.*

Empirical research: Establishing the key parameters for an empirically-validated spatially-unstructured PING model of gamma oscillations in primate visual cortex

Experimental Data

In our empirical data, comprising single-unit and LFP recordings in macaque area V1 (Roberts et al., 2013) and source reconstructed human MEG localized to visual cortex (Hadjipapas et al., 2015), we have observed a robust linear increase in gamma oscillation frequency with increasing luminance contrast. This was consistent across single unit spike trains, LFPs and MEG. Additionally, the slopes of the linear models describing the contrast-to- gamma-frequency relation were similar across this vast range of scales varying from single units to LFP and to MEG. However, there was a difference in the contrast-to-power relationship among measurement scales. In particular, at high grating contrasts, there was a saturation of power or even a robust decay in gamma power in the single unit spike trains and LFP (see Figure 8). However, the MEG power showed no saturation and increased linearly with contrast (Hadjipapas et al., 2015). This is an interesting dissociation, the full understanding of which, will require the use of complex structured models including layers and horizontal connections. Before attempting this, however, the crucial *cross-scale* link between single unit (*micro-*) and LFP (*meso-*) behaviour needs to be investigated. What underlies the frequency shift and the power decay/saturation observed at these two scales? Investigating this question carefully is a prerequisite before one can successfully examine the forward model underling the MEG signal (macro-scale). Validation was based on data from large datasets in three monkeys, two recorded with laminar probes in in area V1, one recorded with a sub-dural ECoG array covering a large part of one hemisphere (Rubehn et al., 2009).

Spatially unstructured V1 model

We developed spatially unstructured and physiologically validated Pyramidal Interneuron Network Gamma (PING) models of excitatory (E) and inhibitory (I) cells to investigate gamma oscillation mechanisms observed empirically in monkey V1 (Figure 16A). Importantly, we developed an approach to calibrate parameters, which are on the one hand largely unknown, but on the other hand are crucial for model behavior. Among the many parameters to be set we focused on the connectivity from E-to-I and from I-to-E cells which has a key role in gamma rhythm generation in PING models (Figure 16B). The calibration was performed by systematically manipulating these parameters and requiring that the model outputs satisfy certain empirical parametric constraints both at the micro-level (contrast-modulated single unit average rates) and at the meso-level (contrast-modulated spectral features of the LFP). The approach was based on a multidimensional fitting of the model against empirical parameters and allows for the selection of valid models, which best reproduce empirically-observed phenomena, including the frequency shift and power decay with increasing contrast and realistic firing rates for E and I neurons (Figure 16C, D). Model parameter fitting also allowed us to make empirically motivated choices in terms of model type (e.g., weak vs strong PING), type of simulated input, and effect of input on E and I cells (Figure 17).

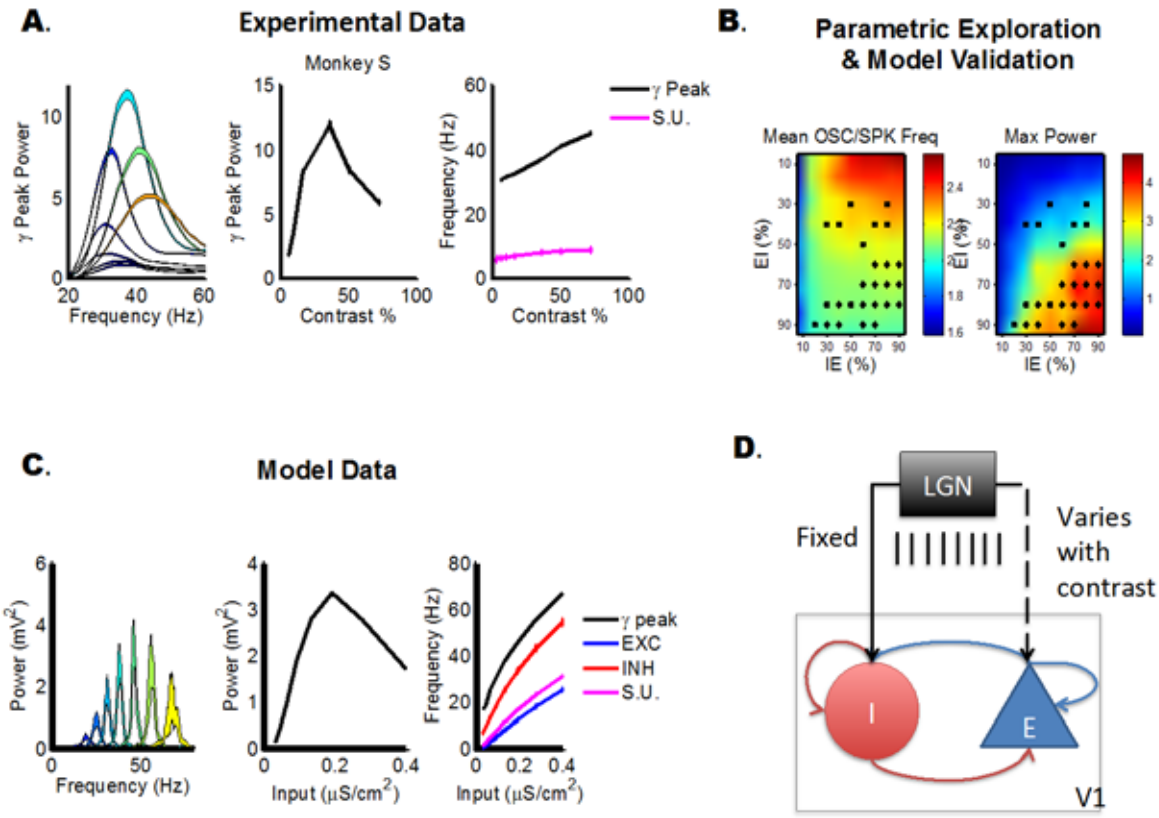


Figure 16: Empirically constrained modelling by parameter fitting

(A) Contrast-dependent modulation of LFP and spikes in macaque V1. V1 LFP spectral response in a representative monkey showing a frequency-dependent spectra shift with contrast (left), a graph extracting from the spectral response the non-monotonic power rise and decay with increased contrast (middle), and a monotonic increase in peak-power gamma oscillation frequency in the LFP (black) and the average single unit firing rate (magenta) (right). Power spectra shown in hotter colours correspond to contrast stimulation, bluer colours correspond to lower contrast conditions. Note the much higher LFP gamma peak frequency representing the population signal (black line) compared to the average single unit firing rate (magenta). This suggests a sparse rhythm in the neural population in which only a subset of neurons participates in the population oscillation in a given cycle. This also is in line with the existence of subthreshold oscillations, and is consistent with the notion of sparse rhythms and weak-PING models **(Wang, 2010)**. **(B) Parametric Exploration and Network Validation.** In order to validate the model, we took a robust set of features from the empirical LFP spectra and the average single-unit firing rates (see Panel A) from our experimental data as well as from the literature. Here, we illustrate a parametric exploration of the model for E-to-I and I-to-E couplings for two out of five of the empirical features selected namely the ratio of the mean frequency of the oscillation versus the average single unit rates (left), and the maximum power observed across all contrasts (right). Colour coding in the two surfaces represents the magnitude of the empirical feature under consideration (i.e., frequency and power). Model output is evaluated across five such empirical criteria and the couplings which satisfy them all are denoted with a black sign. We had monkey V1 recording datasets showing gamma power decay, and others showing gamma power saturation. In the parameter space, valid networks showing saturation are denoted with a diamond symbol, and valid decay-exhibiting networks with a square. The distribution of the two different symbols show different clusters of parameter settings that are required to simulate saturation versus decay behaviour. **(C) Valid PING model.** Example of a valid weak PING model that exhibits a power decay in the LFP spectrum and satisfies all the selection criteria. **(D) Network Diagram of PING network.** The PING network model of randomly coupled regular-spiking excitatory and fast-spiking inhibitory cells modelled as Hodgkin-Huxley-style single compartment neuron models. Coupling within and between populations is random (through voltage-dependent synapses) with a certain probability of connection. The LGN afferent input is modelled in terms of Poisson spike trains. The strength of input to the excitatory cells is varied to represent stimulus contrast-dependent input. For justification of using contrast as an experimental proxy of afferent input see **(Hadjipapas et al., 2015)**

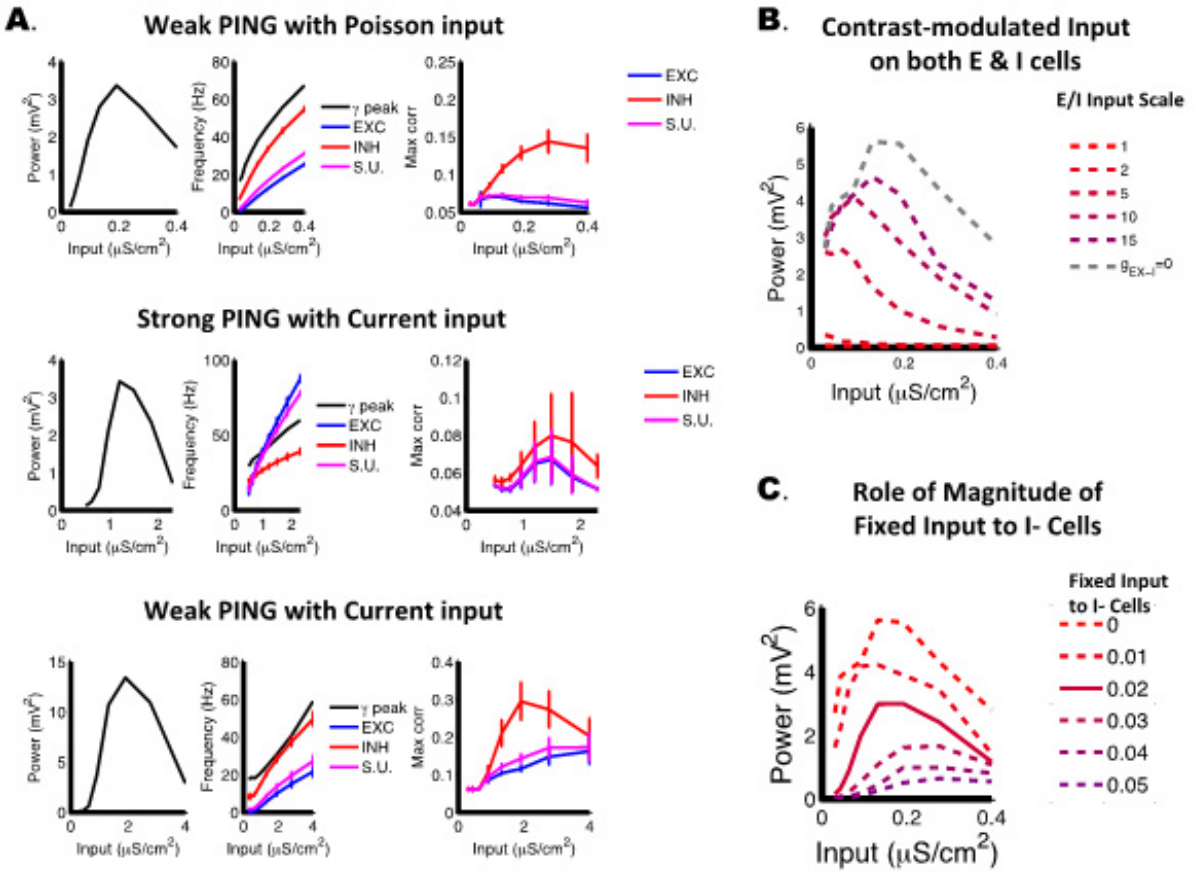


Figure 17: Empirical validation of model choices and type of model input

A selection of the evaluations made is illustrated (A) **Weak and strong PING models.** Vertical columns show parametric variation of network oscillation features as a function of input strength. Left column shows gamma power of modeled LFP. Middle column shows average single unit firing rates for E, I and all neurons (denoted S.U) as well as LFP gamma peak frequency (gamma peak). Right column shows average bivariate correlations among all E (blue), among all I (red) and among all E and I neurons (magenta). Rows represent simulation results from different model types. Top row shows results from **a weak PING model network with Poisson input** as in Figure 16D The increasing synchronization (Max Corr) of the inhibitory cells from low to middle inputs saturates at higher inputs, and shows slight decline at the highest input level used. Middle row shows **strong PING with current input** (used in (Roberts et al., 2013)), which does not satisfy all the criteria although it exhibits the key LFP-based features such as frequency shift and power decay. Bottom row shows **weak PING model with current input**, but modified so that not only the mean but also the standard deviation of the input across neurons varies with contrast, resulting in a valid network with similar decoupling mechanism as in the case of Poisson input (top rows). **(B) Investigation of effects of input to E and I populations: Contrast modulated input on both populations.** When contrast-modulated input was varied simultaneously on both populations, a valid network could not be obtained even when the input to the I-cells was scaled by a factor of 15 versus the input to the E-cells. **(C) Investigation of effects of input to E and I populations: Role of magnitude of fixed input to inhibitory cells.** When running the network simulations for different fixed thalamic inputs onto the I-cells (i.e. using different fixed synaptic conductances), then valid networks were only obtained for intermediate values (as shown in continuous line). Increase of conductance to I-cells destroyed the oscillations evidenced by a marked decrease of power, whereas a decrease of conductance to I-cells altered the oscillation into different frequency bands (consisting of slower, non-gamma frequencies).

Conclusions

Our work, based on empirically validated modeling through multidimensional parameter fitting, and partly illustrated in Figure 16 and Figure 17, has generated the following key findings:

(1) ***Weak PING:*** The likely generative mechanism of gamma falls under the category of weak PING based on a set of validated parameter settings. The main differences of weak (e.g.,

Zachariou et al. 2015) and strong PING models (e.g., Roberts et al. 2013) can be traced back to (a) the synaptic conductances; (b) the E-to-I and I-to-E connection probabilities and; (3) the frequency-input response curve of the excitatory cell model.

(2) ***Micro-scale versus meso-scale divergence:*** The population (LFP) frequency is much higher than individual firing rates, in line with a sparse spiking rhythm in the network, in which only a subset of neurons participates in the population oscillation in a given cycle. This is consistent with the notion of weak PING (Wang, 2010). This result shows a divergence between network behavior at the micro and the meso scale that has important implications in terms of the choice of more detailed generative models that would aim to reproduce the empirical data. This finding supports the necessity of including both spikes (micro-level) and LFP (meso-level) data in modeling.

(3) ***Individual variability*****:** Changes in the specific balance in the strength of E-to-I and I-to-E connectivity within the category of weak PING models permit reproducing variability observed across monkeys in terms of the spectral profile (decay/saturation).

(4) ***The importance of correctly simulating input:*** In order to obtain realistic model behavior, the effects of changes in E-drive (input strength) should affect mostly the E-cells in the model, whereas the effects on I-cells should be limited. In addition, both the mean and the standard deviation of the average input to the E-cells have to be made dependent on contrast.

(5) ***Power decay with increasing contrast:*** The LFP power decay in weak PING model likely results from a decoupling among I-cells at high input strengths (Zachariou et al., 2015)

Empirical constraints for vertically and horizontally expanded V1 models

Constraints for laminar model expansion

Area V1 in the macaque has 6 distinct anatomical layers of between 0.5 and 0.1 mm thickness (Lund, 1988; De Sousa et al., 2010) that are linked by complex patterns of vertical connectivity. In studying gamma processes two functionally distinct domains emerge; the superficial domain above layer 4, and the deep domain below (Maier et al., 2010). Within each domain activity is strongly synchronous whereas synchrony is sharply reduced between the two domains. We therefore divided our data according to these two domains to examine whether empirical constraints defined for the undifferentiated model need be further refined for a two-compartment laminar specific model.

In both monkeys both the superficial and deep LFP gamma generating domains showed similarly shaped gamma power contrast response functions and equal amounts of suppression/saturation, despite small yet significant differences in the level of gamma power and spike rates. There were, moreover, no differences in gamma peak frequency between laminar compartments despite significant differences in firing rates. This implies a different level of sparseness in the two domains. These observations are consistent with weak PING generative mechanisms in the two domains, and with only minor adjustments in the model settings required to account for different sparseness levels. Computational models of superficial versus deep laminar differences will thus be highly similar.

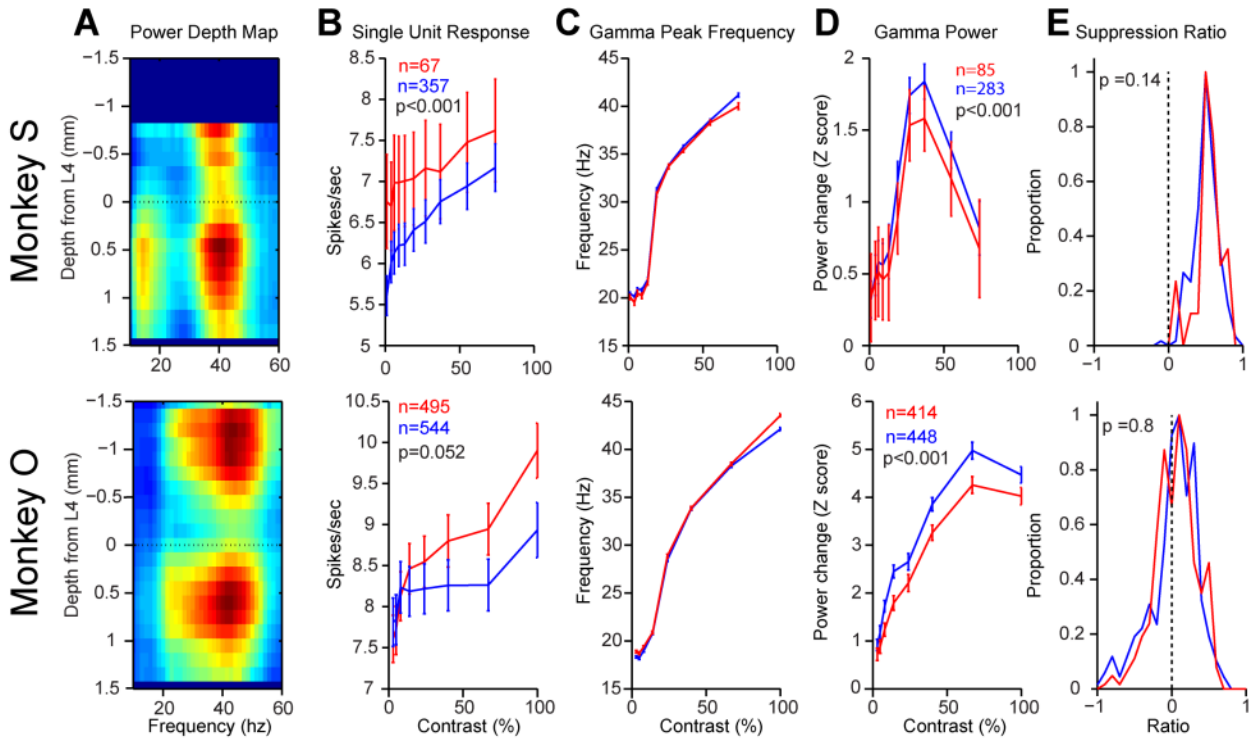


Figure 18: Laminar differences in constraint parameters

Upper row shows data from monkey S; lower row data from monkey O. **(A) Laminar distribution of power spectra** (change in power from baseline during high contrast stimulation). Y-axis shows the depth of the electrode's recording contact relative to the depth of layer 4; negative values indicate more superficial positions. Red colors indicate an increase in power from baseline. Notice two regions of high power increase centered around 40Hz corresponding to the gamma domain in the superficial and deep layers. Dashed line at depth zero indicates the level of layer 4 and where the two domains were separated for analysis. In Monkey S, superficial layers seemed thin, and were under sampled. **(B) Mean single unit spike rates during the sustained period of stimulation** (from 250ms after stimulus onset) in superficial (red) and deep (blue) domains as a function of contrast. Error bars show ± SEM. N values indicate the number of single units recorded in each domain, ANOVA P-values indicate significance of the difference between spike rates in each domain. Note that the P-value in monkey O is marginal probably because the responses were overlapping at very low contrast ranges, however in both monkeys, differences in firing rates were subtle. **(C) Gamma band peak frequency as a function of stimulus frequency**. Lines show means, error bars show ± SEM. Red lines correspond to superficial domain, blue to deep domain. Number of observations are as in D). Differences are non-significant. **D) Gamma peak power as a function of stimulus contrast**. Peak power was computed as Z-score of change from baseline. Lines show means, error bars show ± SEM. Red lines correspond to superficial domain, blue to deep domain. Notice the suppression of gamma power at high contrasts. In both monkeys, gamma power was marginally higher in deep rather than superficial layers, contrary to other literature (Maier et al., 2010; Roberts et al., 2013). This may reflect differences in analysis or the chosen boundary – note the 0.3 mm of low power above the zero line, or other factors. **(E) Quantification of power suppression at high contrast**. The strength of suppression per contact was quantified as the change in power from the preferred contrast to the highest contrast (37% in monkey S, 67% in monkey O), divided by the power at the preferred contrast. Values below 0 indicate suppression, values above zero indicate higher power at the highest contrast. Lines show histograms of observed suppression ratios. P-values indicate significance (two-sample t-test) of differences in suppression ratios between superficial and deep gamma domains. We found no significant difference in the suppression ratio between the superficial and deep domains.

Constraints for horizontal expansion of laminar models

The MEG signal is a global measure of brain activity. As it depends on the aggregation of the activity of a wide volume of cortex, highly synchronous activity in these volumes may be

expected to give a higher-amplitude MEG signal (Hadjipapas et al., 2015; Hämäläinen et al., 1993; Nunez and Srinivasan, 2006b). We therefore reasoned that the contrast response function of MEG gamma would depend on the correlated activity within and across domains. We found recently that while LFP (and single unit) gamma power showed a marked nonlinearity, decreasing or saturating at high contrasts, in the MEG data this was not the case; gamma power scaled linearly with contrast (Hadjipapas et al., 2015). This disassociation between findings might be accounted for by the differences in the lateral extent of the signal generator (and lateral synchronization), since the source generators of aggregate macroscopic signals such as MEG and EEG presumably depend not only on local network power but also on larger scale lateral synchronization between such local networks (Musall et al., 2014). We thus reasoned that it is possible that power increases of the aggregate MEG could be reconciled with a concomitant decrease/saturation of power in local network signals (as measured by LFP and subdural ECoG) if contrast would increase the lateral synchronization between these local networks. Under this hypothesis, the lateral (spatial) synchronization contribution to the aggregate signal would outweigh a decrease in the local signal power component (Musall et al., 2014). We first tested this idea by calculating a signal composed of the multiple LFP and ECoG channels averaged in the time domain (aggregated). We tested aggregation of channels arranged horizontally over the cortical surface, using data from surface ECoG (acquired during the same paradigm), and channels arranged vertically using data from laminar probes. We calculated the contrast response function of gamma in these aggregated signals, and found equal suppression or saturation in these signal as in the raw LFP and EcoG signals. This suggests that synchrony between remote neural populations did not increase with contrast. To test this further, we calculated the correlation in the time domain of LFP and EcoG data recorded at separate electrodes, at zero millisecond time lag (zero-lag correlation) after filtering LFPs in the gamma range. We found that zero-lag correlations were significantly reduced at high contrast. This analysis indicates that, contrary to our original hypothesis, synchrony between horizontally connected model neuron networks should decrease with increasing contrast.

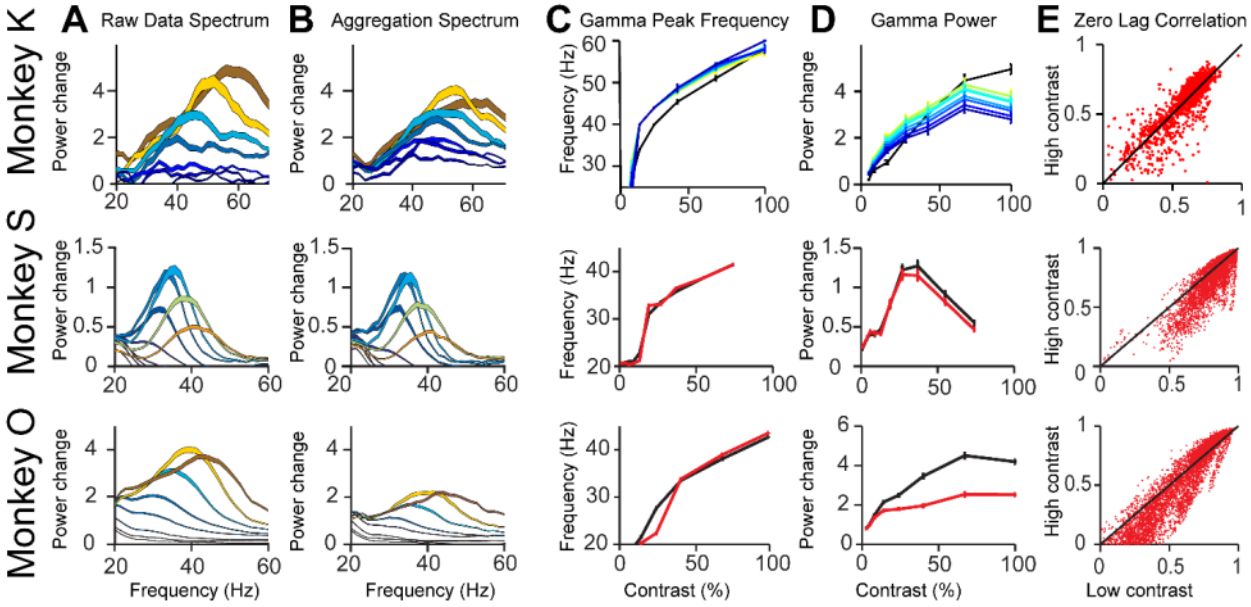


Figure 19: Effects of horizontal and vertical signal aggregation.

Data from different rows correspond to different monkeys, columns to different analyses. First row shows analyses of horizontal aggregation on monkey K, which was recorded with surface ECoG covering most of one hemisphere. Data analysis was restricted to electrodes over V1. Second and third rows show data from monkeys S and O recorded from with laminar probes in V1. **(A) Change in power spectrum from baseline in raw LFP**. Hotter colors correspond to power spectrum during higher contrast stimulation, bluer colors correspond to lower contrast stimulation. Line width indicates ±SEM. **(B) Change in power spectrum in aggregation**, conventions as in A. For Monkeys S and O vertical aggregation was over all V1 electrodes of one laminar probe. For Monkey K we tested lateral aggregation over

distances from 2mm to 2cm (electrode spacing=2mms). Data shown are for aggregation over 6mm. **(C) Contrast response functions of gamma peak frequency.** Contrast response gamma frequency was computed for each individual raw LFP (black) and aggregation data (red in monkeys S and O). In monkey K, yellow indicates aggregation over 0.2 mm, to dark blue indicates aggregation over 2cm) **(D) Contrast response function of Gamma power**, line colors as in C. **E) Change in zero-lag correlation between high contrast stimulation and low contrast stimulation**. Each dot shows values from one contact pair. Points below the diagonal correspond to a reduction in correlation with increased contrast. Significance was tested with paired t-test, P values were <0.001 in all cases.

Laminar Signal derivation

The MEG signal is sensitive to magnetic fields arising from current flow along long dendrites which run perpendicular to the cortical surface (Hämäläinen et al., 1993). Of particular interest are the pyramidal cells with a soma in layer 5 and dendrites reaching in layers 1 and 2 because these neurons have the longest vertical dendrites they may generate the largest magnetic fields(Hadjipapas et al., 2015; Lee and Jones, 2013; Murakami and Okada, 2006). We hoped to isolate the activity of large neurons with trans-laminar arrangements by calculating the difference between LFPs recorded at different cortical depths.

We first examined the laminar structure of zero lag correlations between LFPs (filtered in the gamma range) recorded at different depths. Consistent with previous literature (Maier et al., 2010) channels within the upper and lower domains were highly correlated (red regions in upper left and lower right of Figure 20A) whereas correlation between the domains was much weaker (green/blue region in upper right). We tested the effect of changing stimulus contrast on the correlation between laminar domains, and found that zero lag correlations were reduced at higher contrast (Figure 20B and C, also see Figure 19E). Interestingly, correlations between domains (Figure 20C, green line) dropped more precipitately than correlations within domains (red and green lines). Correlations within the superficial domain (Figure 20C, red line) dropped the least with increasing contrast. As we had hypothesized above, contrast related changes in the zero-lag correlation would lead to differences in the gamma power contrast-response function. Specifically, we found that for the signals derived from superficial-to-deep differences (green line, panel E) and from within-superficial channels (red line, panel E) there appeared to be *less* suppression of gamma power at high contrasts, as compared to the raw LFP (Black line). Signals derived from pairs of channels both within the deep domain (blue line) appeared to show *stronger* suppression. In both monkeys there was a significant difference in suppression ratio between the groups (ANOVA, $p<0.01$*. Post-hoc testing showed that in* monkey S all derivation classes were significantly different from each other, with derivations between pairs of deep channels showing the strongest suppression, and derivations between pairs of superficial channels showing significantly weaker suppression. In monkey O, derivations between pairs of superficial channels displayed significantly reduced suppression and showed a trend towards negative suppression ratios (i.e. increased power at the highest contrast). Derivations between superficial channels were significantly different from the other two classes (which did not differ significantly from each other).

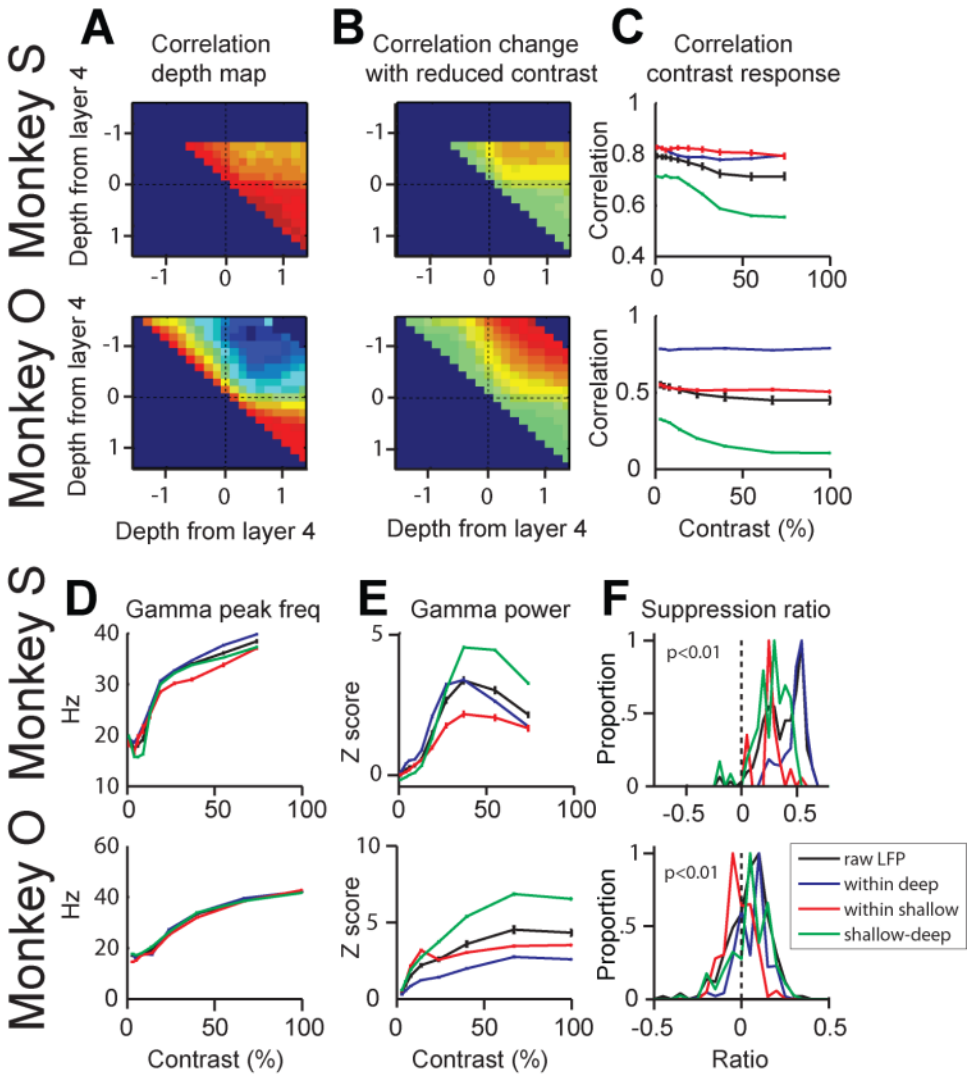


Figure 20: Laminar derivation analysis

(A) zero-lag correlation values between pairs of LFP channels at different depths **Regions in the upper left quadrant show pairs where both channels are above layer 4, regions in the lower right show correlations between pairs where both channels are below layer 4. Regions in the top right quadrant show pairs which cross the layer 4 boundary. Red colors indicate strong correlations blue indicate weak correlation; consistent with previous literature (Maier et al., 2010) the layer 4 boundary is marked by a sharp reduction in correlations across the boundary, whereas correlations with either gamma domain are high.** (B) Change in zero-lag correlations with reduced contrast **(difference between the highest contrast and the lowest). Red colors indicate an increase in correlation with reduced contrast, green indicates little change. A novel finding is that the gamma domains become more independent with increasing contrast.** (C) Subdivision of data: **To better illustrate the contrast dependence of zero-lag correlations data were divided into four categories: the raw LFP is shown in black, correlations between pairs of superficial channels is shown in red, between pairs of deep channels in blue and pairs crossing the layer 4 boundary are shown in green. Zero lag correlation values are shown as a function of contrast for each class, error bars show ±SEM. Notice that the green curve drops more dramatically with increasing contrast.** (D) Further power analysis: **We calculated the power spectrum of signals created by taking the difference between pairs of LFP channels at different depths. Data is grouped as in C. All pairs showed the same changing peak gamma frequency as a function of contrast.** (E) Gamma peak power as a function of contrast for separate LFP-difference classes. **Unsurprisingly the differences between superficial and deep channels (green) showed the highest power since, taking the difference between highly correlated channels will result in low amplitudes. Of interest is the change in shape of the gamma power-contrast response functions between the different classes. Green and red curves appear to show reduced suppression at high contrasts.** F) Suppression/saturation **was again quantified by the suppression ratio as in Figure 12E and compared across classes. Curves show histograms of suppression ratios for each class (line color). Positive values indicate strong suppression while positive values indicate increased power at the highest contrast. In both monkeys the four classes showed significantly different suppression ratios (p values in figure give the significance from an ANOVA).**

Conclusions

Our new analysis of the contrast dependency of gamma band activity and correlations in area V1 revealed several novel insights.

(1) ***Weak PING in superficial and deep layers***. The ratio of firing rate to gamma frequency supported a weak-PING mechanism and that a similar, albeit not identical, degree of sparseness existed in the deep and superficial gamma generating domains. Our large-scale model will be built of several coupled weak-PING networks. The strength of coupling between the networks will determine how independently each network operates.

(2) ***Zero-lag correlations are reduced at high contrast***. These observations will be useful as constraints for the contrast response function of vertical and horizontal model coupling strengths. The lack of saturation in MEG power at high contrast does not appear to be due to increased synchrony either within or between cortical columns.

(3) ***The contrast-response function of gamma in specific trans-laminar potential differences offers interesting avenues for future work to understand the generative mechanisms of MEG***. We hypothesized that *signals derived from the combination of local network activities would have a different contrast-response function than that observed within the local networks, if the coupling between the networks changed as a function of contrast*. In line with this, we found that gamma power in the signal derived from the derivation of LFPs at specific laminar depths showed less suppression at high contrast. We anticipated that the largest shift in the gamma-band-power contrast-response function would be present in the deep-superficial derivation as we conjectured that this would best approximate the current dipole thought to underlie MEG. However, this was not supported by the data. We found weak support of our hypothesis in monkey S, where the deep-superficial data showed significantly less suppression than either the LFP or the within-deep data. However, in monkey O, where more data was collected in the superficial domain, is was clear that only the within-superficial derivation showed less suppression at high contrast. Notably, in Monkey S the within-superficial derivation also showed the weakest suppression. Further work will be required to fully understand this observation and, especially to understand whether this observation is relevant to understanding the generative mechanisms of gamma observed in human MEG.

A proof-of-principle layer-extended column model

Empirically informed models are important for realistic perceptual processing and realistic interactions with high-level cognition. However, empirical validation of large scale models with extensive parameter space is challenging and can benefit from constrains from experimental data across different modalities (e.g. spikes, LFP and MEG).

Our experimental analysis has armed us with important insights on how to progress in developing an empirically-constrained, layered, cortical-column model informed by laminar analysis of LFPs and unit spikes in monkeys and to further compare this with MEG recordings in humans recorded during the same experimental protocol. The role of the pyramidal cells in V1 layer 5 has been suggested previously as a key component in the current dipole observed in MEG signal (Hadjipapas et al., 2015; Jones et al., 2009; Lee and Jones, 2013). Moreover, our trans-laminar analysis indicated that that differential LFPs within the superficial domain and across superficial to deep contacts reveals a shift in gamma power contrast response functions. This signal could be most sensitive to the activity of trans-laminar Pyramidal cells as the MEG is also considered to be.

Hence, as a first proof-of concept approach we considered a well studied integrate and fire network (Brunel, 2000) which is analytically tractable and has been shown to exhibit a variety of dynamical states such as asynchronous irregular and fast or slow oscillations with irregular spiking activity. This network is incorporated in the hybrid LFPy, a recently developed hybrid framework which forms the basis for LFP generation when input is received from point-neuron network models (i.e. LFP is generated by multicompartmental model neurons and can be 'read off' at various electrode positions following a specification of a detailed forward model, as illustrated in Figure 21A), such as the Brunel network (Hagen et al., 2015). We simulated this network in a regime where it exhibits oscillations in the gamma range and modified/increased the input such that an asynchronous irregular activity regime was obtained. We observed (see Figure 21B) that the contrast-response function was similar across all depths with quantitative rather than qualitative differences, in line with findings above (Figure 18). The model showcases the value of the ground truth CSD data, that can be directly obtained in this framework. The CSD localizes the activity (the underlying transmembrane currents) to the correct contact locations/depths 5 (this is where the neuronal somata were placed - see panel A) and 2-3 (this is where many synapses were made). This can be seen in the bottom plot in Figure 21, where only these 3 contacts show significant power. This is to be contrasted with the LFP, where this localization is smeared out with many of the other contacts also showing gamma power, which is simply due to volume conduction. Most importantly however, this model also reproduced some of the key features of the empirical data such as a gamma peak frequency increase with input and a nonlinear gamma power modulation (power saturation) at high input. In addition, the LFP oscillation frequency was higher than the average E and I unit firing rates, although this difference was small compared to the real data and the previous spatially- undifferentiated weak- PING model. Related to this, the expected difference in E- to I- cell populations average firing rates was not observed. This was however expected, as by construction this specific model network was comprised of identical cells models for both populations. Nonetheless, this study shows that the hybrid- model framework can be useful in further exploring laminar differences and interactions. This can be achieved by extending and validating this network such as to optimally reflect our experimental findings, in an analogous fashion to the work performed in the undifferentiated model case. This is will be further pursued in collaboration with the groups of Gaute Einevoll (SP4) and Markus Diesmann (SP6).

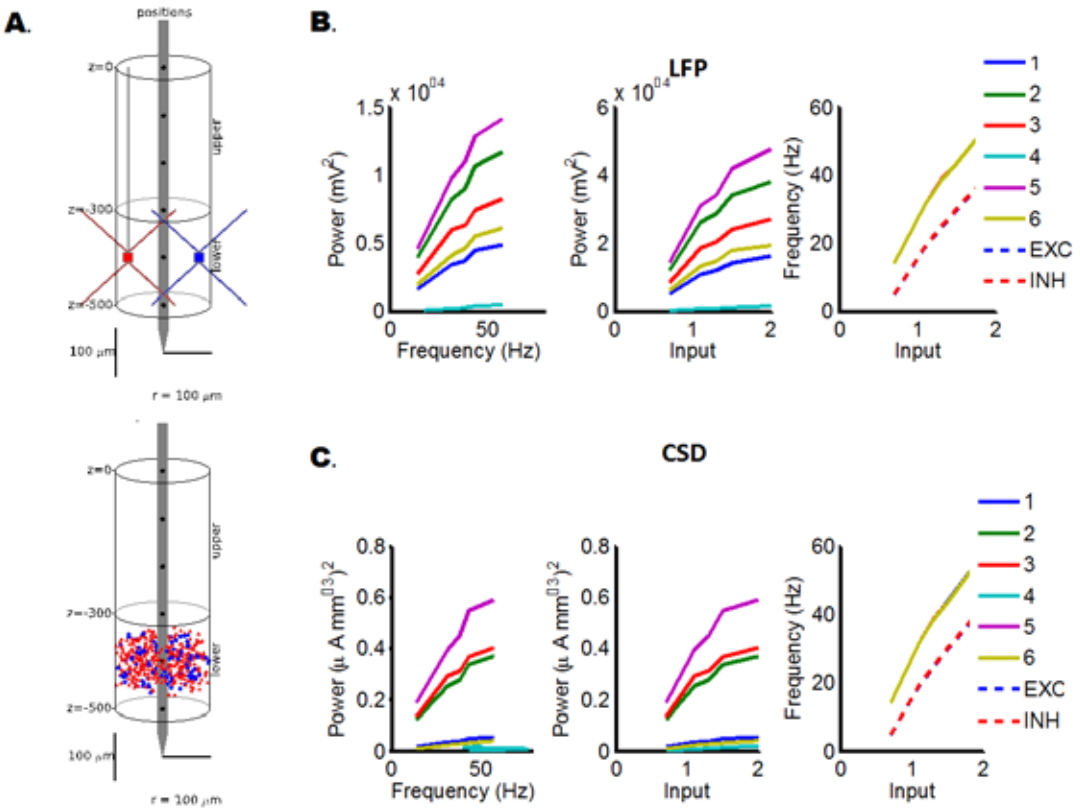


Figure 21: A Layer-extended column model

We used an existing model (Brunel, 2000) and modified it to approach empirical data from monkey recording experiments (Roberts et al., 2013). **(A) Structure of the model.** A schematic view of the morphologies (top) and the location of the somas (bottom) of the E and I cells and the location of the 6 recording contacts across the two layers, deep and the superficial. **(B) Spectral characteristics of the LFP.** LFP data are shown for each of the six recording contacts across trials for various Inputs (contrast), and average firing rates of the E and I populations. The power of channel 4 was too low to estimate frequency properly and hence it is not plotted in the left panel. **(C) Spectral characteristics of the CSD.** Spectral CSD data is shown for each of the six recording points across all for the three inputs (contrast) low, medium and high, and average firing rates of the E and I populations. The power of channel 4 was too low to estimate frequency properly and hence it is not plotted in the left panel.

General Conclusions

Our work has led to various novel insights both from experimental data analysis and models as listed below.

Experimental data analysis:

1. In single-unit and LFP recordings in macaque area V1 (undifferentiated across all layers) and source reconstructed human MEG localized to visual cortex we have observed a robust increase in both gamma oscillation frequency and single unit spiking activity with increasing luminance contrast. In addition, at high grating contrasts, a robust decay in gamma power was observed in the LFP but not the MEG.
2. Both the superficial and deep gamma-generating laminar domains showed similarly shaped gamma power contrast response functions and equal amounts of suppression/saturation, despite small yet significant differences in the level of gamma power and spike rates (sparseness).

3. Horizontal LFP aggregation does not shift the gamma power contrast response functions (zero lag correlation stays the same and drops with increasing contrast) contrary to our original hypothesis.
4. Differential LFPs across superficial and across superficial-deep contacts reveals a shift in gamma power contrast response functions. This signal could be most sensitive to the activity of those trans-laminar pyramidal cells the MEG is thought to be most dependent upon. Further work will be needed to better understand these findings.

Modeling:

1. Out of many possible models, the empirically validated ones pointed to weak-PING (vs strong PING) model as the likely gamma mechanism.
2. A specific balance between the strength of E-to-I and I-to-E connectivity is required for realistic network behavior and to reproduce certain variability observed across monkeys in terms of the spectral profile (decay/saturation).
3. In order to obtain realistic model behavior, the effects of changes in E-drive (input strength) should affect mostly the E-cells in the model, whereas the effects on I-cells should be limited. In addition, both the mean and the standard deviation of the average input to the E-cells must be modulated by contrast.
4. The LFP power decay in weak PING model likely results from a decoupling among I-cells at high input strengths.
5. An appropriate framework with biophysical LFP calculations and laminar cell morphology is essential when studying laminar LFP results.

Location of our data storage:
To be assigned weblink on UM server (see Dataset Information Card Task 3.1.4 "Models of gamma oscillations in visual cortex")

Provenance of the data:
Modeling data (see DIC Task 3.1.4 "Models of gamma oscillations in visual cortex")

Self-analysis of the value and completeness of our data

- We have shown that there are striking gamma spectral changes as a function of contrast in monkeys and humans. Moreover, we have shown that while the human macro-level MEG signal shares many features with the invasive LFP and spiking signals, there are differences in the contrast-response function of gamma power. These spectral responses to contrast (E-drive) are informative with respect to structure and function of models. Our aim was to do complementary modeling and empirical data analysis to build empirically validated PING models, and then to use these validated PING models as building blocks laterally and vertically expanded models of primary visual cortex. In the modeling, we aimed to compare local gamma generating mechanisms, observable only in invasive recordings in animals, with the global gamma generating mechanism, which is observable in humans using MEG. Our ***hypothesis*** was that the invasive signal is sensitive to local processes, while the MEG is sensitive to global features, including synchronization and time delays between local networks (and we tested several specific hypothesis yielding surprising and novel insights, see Results). Because our modeling is empirically constrained a large effort was done in terms of empirical analysis of MEG and especially monkey spiking and LFP data recorded in the same contrast manipulation paradigm prior to the onset of the present project. The ultimate goal was the

development of a computational model that would faithfully reproduce the behaviour of the invasive data at the local level and of the MEG data at the global level.

- Developing such a model faces several challenges. Chief among them is that there are many free model parameters which are unknown and so often set by convention without empirical support. To overcome this we developed ***a novel method for validating model behaviour against empirical data***. As part of this validation process, we performed novel analysis of our empirical monkey datasets to produce previously sparsely-reported descriptive statistics of electrophysiological parameters in macaque V1. These ***novel statistics*** will be made available through the HBP platform and we predict they will be beneficial to other modeling groups. We have fully developed an ***empirically constrained unstructured model of gamma in V1*** that has already produced novel insights.
- To achieve the transfer function to MEG requires developing the model to a structured network including laminar and columnar structure, constrained by empirical observations of dependencies and time delays between these components. Our analysis of invasive data sets has produced ***novel insights in horizontal and vertical (columnar) gamma generating mechanisms***, which will point the way for further model development. The data are informative on the manner in which both the vertical extension (columns) and lateral extension can be constrained. We have started working with a prototype columnar model previously published, and started modifying it using empirical constraints in a manner applied to our unstructured PING model.
- Our involvement in the HBP has allowed us to develop ***collaborations*** with other groups, who will continue in the HBP, with whom we hope to further develop our model.
- ***In summary***, our project has produced completed empirical and modeling work on unstructured PING relevant to KPI_1; significantly advanced empirical analysis and ongoing modeling work relevant for KPI_2; and highly relevant empirical analysis that will constrain models in KPI_3 and 4 (which are two KPIs that are in fact closely linked). Hence, we have delivered significant progress on all aspects of the project, with part of the work completed. The work has yielded highly valuable empirical constraints, novel methodology for empirically constrained modeling, novel statistics and novel insights into the generative mechanisms of gamma in primary visual cortex. Parts of the project are already published; further work is at an advanced state of preparation and will be submitted in the coming months. Our modeling work will continue through collaborations we have developed through our involvement in HBP (with the groups of Diesmann and Einevoll). We regard this as an *excellent level of output*, especially given the imposed limits in time and funding.

Scientific output and data use

Review

Unpublished, included in Deliverable report.

Publications

Hadjipapas, A., Lowet, E., Roberts, M. J., Peter, A., & De Weerd, P. (2015). Parametric variation of gamma frequency and power with luminance contrast: A comparative study of human MEG and monkey LFP and spike responses. *NeuroImage*, *112*, 327–340. http://doi.org/10.1016/j.neuroimage.2015.02.062

Zachariou, M., Roberts, M., Lowet, E., de Weerd, P., & Hadjipapas, A. (2015). Contrast-dependent modulation of gamma rhythm in v1: a network model. *BMC Neuroscience*, *16*(Suppl 1), O10. http://doi.org/10.1186/1471-2202-16-S1-O10

Publications in advanced preparation based on results in present delivery report

- Empirically constrained PING modelling

- Laminar characterization of gamma responses in monkey V1

Conferences

Zachariou et al, 2015. 24th Annual Computational Neuroscience Meeting (CNS), Prague, Czech Republic, 18–23rd July (BMC Neuroscience)

Zachariou at 2015. 19th Conference of the European Society for Cognitive Psychology (ESCOP), September 17-20th, 2015.

Hadjipapas et al, (spotlight session) and Zachariou et al, (poster), HBP Summit 2015, Madrid.

Ongoing/future colloborations

With groups of Diesmann (SP6 T6.2.3) and Einevoll (SP4 T4.1.2)

1.3 Visual attention and the mechanisms of inter-areal communication

Task T3.1.1 - Pascal Fries (ESI), Chris Lewis (ESI)

Introduction

The greatest proportion of brain activity is endogenously generated. The brain's endogenous activity is highly structured and affects sensory coding, behaviour, and perception. The observation of structured endogenous activity across spatial scales suggests that it plays a role in the maintenance and formation of brain networks. The correlation of spontaneous functional MRI signals has demonstrated the existence of multiple intrinsic networks, previously observed during controlled cognitive paradigms. The prevalence and reliability of intrinsic networks have generated intense interest in the functional relevance and electrophysiological basis of inter-areal correlations. Using multisite recordings from areas V1 and V4 of awake monkeys we investigated the spatial and temporal structure of intrinsically driven activity during both stimulation and passive fixation.

Review of the cortical architectures underlying attentional selection and the communication-through-coherence hypothesis

Pascal Fries "Rhythms for Cognition: Communication through Coherence", Neuron, Volume 88, Issue 1, p220–235, 7 October 2015

Abstract

I propose that synchronization affects communication between neuronal groups. Gamma-band (30-90 Hz) synchronization modulates excitation rapidly enough so it escapes the following inhibition and activates postsynaptic neurons effectively. Synchronization also ensures that a presynaptic activation pattern arrives at postsynaptic neurons in a temporally coordinated manner. At a postsynaptic neuron, multiple presynaptic groups converge, e.g. representing different stimuli. If a stimulus is selected by attention, its neuronal representation shows stronger and higher-frequency gamma-band synchronization. Thereby, the attended stimulus representation selectively entrains postsynaptic neurons. The entrainment creates sequences of short excitation and longer inhibition that are coordinated between pre- and postsynaptic groups to transmit the attended representation and shut out competing inputs. The predominantly bottom-up directed gamma-band influences are controlled by predominantly top-down directed alpha-beta band (8-20 Hz) influences. Attention itself samples stimuli at a 7-8 Hz theta rhythm. Thus, several rhythms and their interplay render neuronal communication effective, precise and selective.

Data set: Large-scale recordings from distributed and local visual networks during rest

Empirical data

We collected data from a custom designed, high-density electrocorticography array covering large portions of the superficial cortex in two macaque monkeys (Figure 22A). We presented a series of visual stimuli in order to map the spatial selectivity of the recorded visual regions (Figure 22B). Additionally, we collected data during periods of passive fixation while the animals awaited a visual attention task. We further investigated the extent to which intrinsic signals relate to the coding of visual stimuli across a broad range of contexts and stimulus classes. Finally, we investigated measures to quantify the reliability of individual neurons, the tuning similarity of pairs of neurons and the intrinsically driven variability of pairs of neurons.

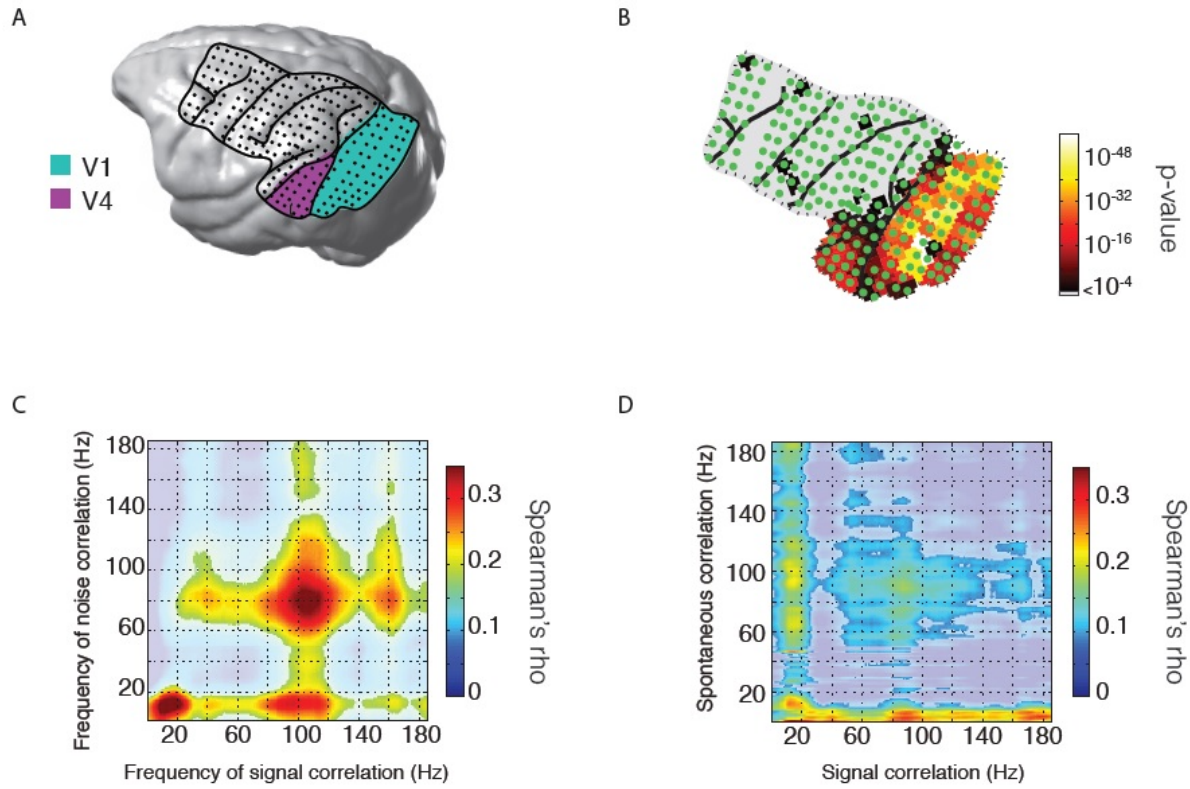


Figure 22: Intrinsic visual activity is topographically organized in local and inter-areal rhythmic synchronization

A) Custom, high-density ECoG array shown on cortical surface model of macaque monkey. B) Visual selectivity of all recorded ECoG channels is limited to recording sites in visual areas. C) Trial to trial variability is structured in both space and time. D) During passive fixation, visual areas are intrinsically coupled in retinotopically specific, local and inter-areal synchronization.

Results

We found that endogenously generated activity, both during stimulation and passive fixation, exhibits a similar pattern of local and inter-areal rhythmic synchronization (Figure 22C and D). Further, we found that these patterns of synchronization were related to the topographical organization of the recorded visual areas. Rather than occurring across a broad range of frequencies, the intrinsic activity was exhibited topographically specific inter-areal coupling at

specific frequencies (Lewis et al., 2016). We next investigated the pattern of activity in visual cortex that contained the greatest amount of stimulus information. We found that the frequencies showing the highest topographically specific organization also contained the most information about visual stimuli (Figure 23). These frequencies contained information about multiple visual stimulus attributes, including: position (Figure 23A), natural image identity (Figure 23B), as well as the visual orientation (Figure 23C). Importantly, the rhythmic activity contained comparable information about grating orientation as the simultaneously recorded spiking activity (Figure 23D).

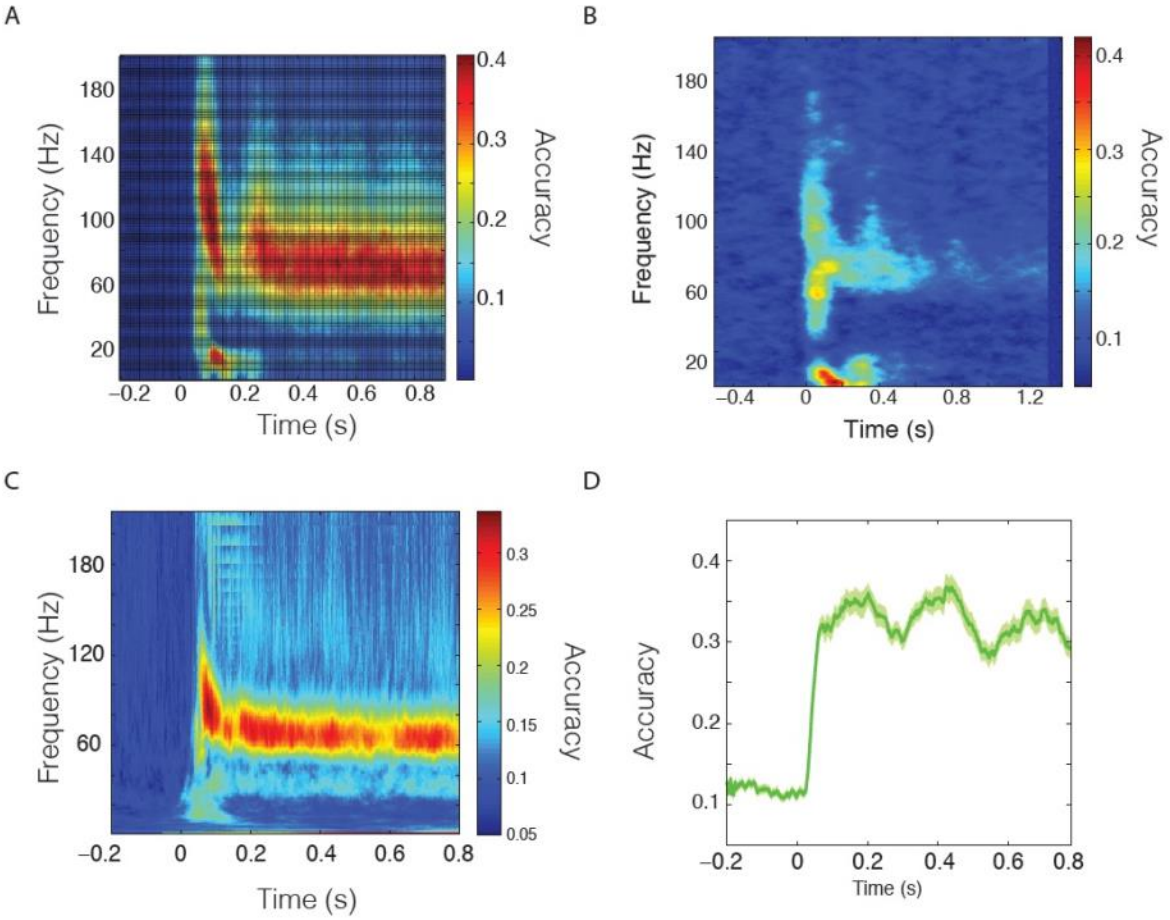


Figure 23: Stimulus selectivity of visual rhythmic activity

A) Information about the position of a visual stimulus is highest in two frequency bands with different temporal dynamics. B) The same frequency bands with stimulus position information have the highest information about the identity of natural visual scenes. C) These frequency bands also contain the highest amount of information about the orientation of visual grating stimuli. D) Information about orientation in visual rhythmic activity is comparable to the information in simultaneously recorded multi-unit activity.

In order to better understand how intrinsic activity effects the responses of single visual cells, we introduced new methods to access the reliability of single cells, as well as the tuning similarity of pairs of single cells (Figure 24).

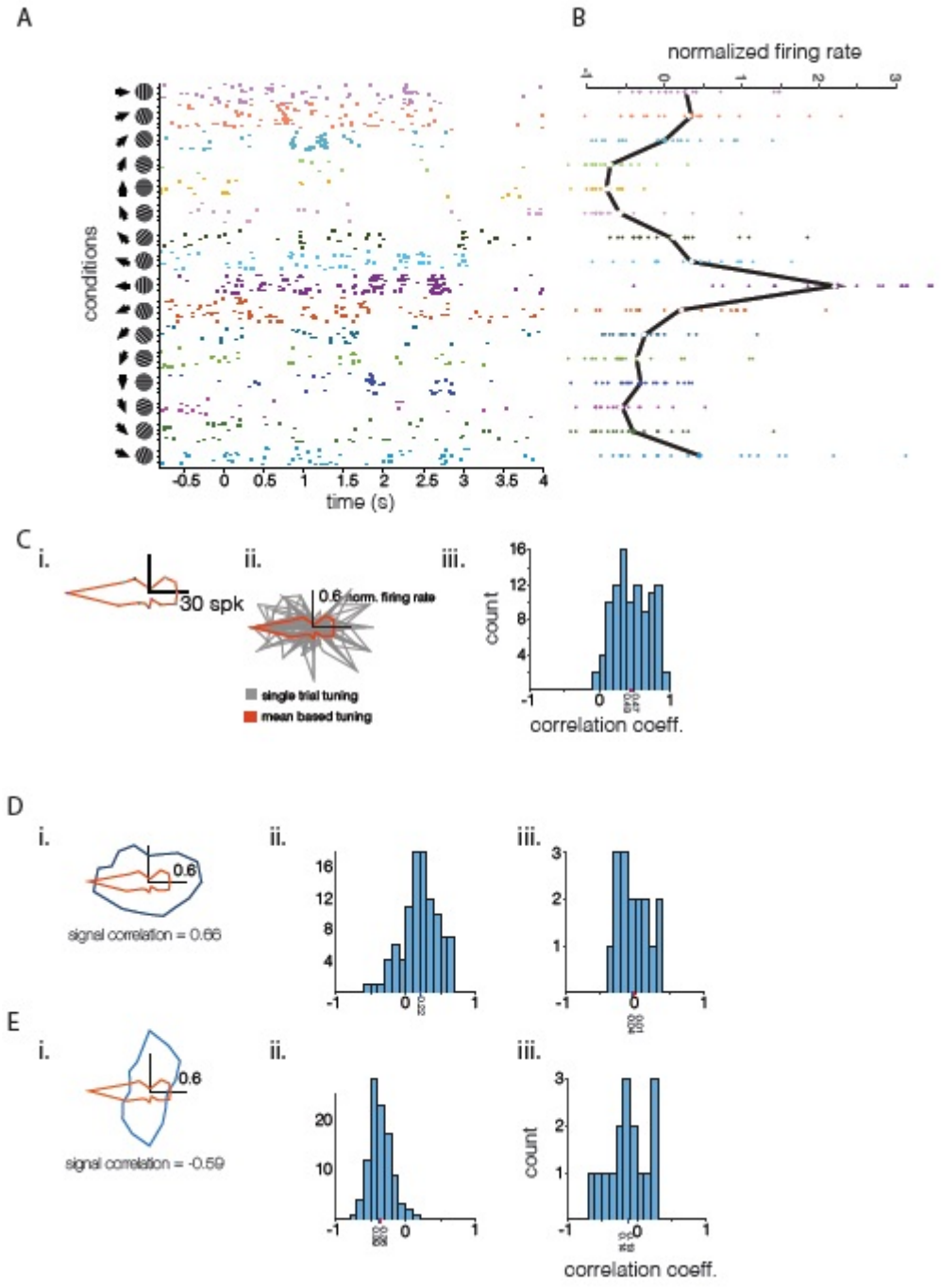


Figure 24: Quantifying single unit reliability, signal correlation and intrinsic correlation

A) The single trial responses of a single V1 neuron to multiple repetitions of differently oriented gratings. B) The mean response (tuning) of the cell shown in (A) to different orientations (in black) with individual trials shown in different colors. C) The reliability of a single V1 neuron. (i) mean-based tuning, (ii) mean-based tuning (in red) overlaid with individual trial estimates of tuning. (iii) single-trial estimates of tuning reliability. D) The tuning single-trial tuning similarity of two similarly tuned cells. (i) Mean-based tuning of cell 1 (in red) and cell 2 (in blue). (ii) single-trial based tuning similarity of the two cells. (iii) noise correlation of the two cells. E) The tuning single-trial tuning similarity of two dissimilarly tuned cells. (i) Mean-based tuning of cell 1 (in red) and cell 2 (in blue). (ii) single-trial based tuning similarity of the two cells. (iii) noise correlation of the two cells.

We first assessed the variability of single cell responses to repetitions of identical gratings (Figure 24A). We found that though the mean response across identical trial repetitions

exhibited the well-known orientation tuning of single V1 cells, there was a high degree of trial variability arising from uncontrolled intrinsic activity (Figure 24B). In order to assess how this intrinsic variability affected the reliability of single neurons, we computed measures of single-trial tuning (Figure 24C), which take into account the individual trial response. These measures indicate that although an individual cells tuning is relatively well preserved, assessing individual trial responses suggests that the responses of V1 neurons relay intrinsic and extrinsic signals roughly equally. We next sought to use these measures to assess the tuning similarity of pairs of V1 neurons (Figure 24D and E). We found that our single trial measures of tuning similarity quantified the extent to which two cells reliably exhibited similar tuning. Because our measures were generated from single trial estimates, we were able to generate populations of values, rather than the single values often used and to thereby get a complete picture of the degree to which pairs of V1 cells share similar extrinsic (tuning) and intrinsic information.

Conclusions

In total, our results combine to suggest that both the stimulus-driven and the intrinsic activity of visual areas are intricately structured in both spatial and temporal dimensions. We found that intrinsic activity, both as trial-to-trial variation in stimulus response to identical visual stimulation, and as spontaneous activity during passive fixation, reflect the topographical organization of the underlying cortex. Further, the topographically specific reactivation occurs in specific frequency bands. Importantly, these frequency bands also contain the most stimulus related information in visual areas across a broad range of behavioural conditions and stimulus attributes. Finally, by better assessing the degree to which individual V1 cells reflect external and internal variables, we can begin to assess how single cells and populations of cells combine extrinsic and intrinsic factors in cortical computations.

A **Dataset Card Information** has been completed (See DIC Task T3.1.1 “Spontaneous activity in anesthetized cat area 17”).

Data Provenance
The data were collected by Conrado BOSMAN and Christopher LEWIS at The Donders Institute for Brain, Cognition, and Behaviour, at Radboud University, the Netherlands, and the ESI.

The location of data
Data are hosted by the Ernst Strüngmann Institute (ESI) for Neuroscience and are available to collaborators upon request. The data is complete and satisfies the commitments made in the description of work.

Collaborations and data usage
Ongoing collaboration on modeling inter-areal dynamics with Gustavo Deco in SP 2. He has used data collected by the ESI in order to validate models of spontaneous activity and to improve methods of source-localization for electrocorticography data.

Publications related to HBP
Lewis CM, Bosman CA, Womelsdorf T, Fries P (2016) Stimulus-induced visual cortical networks are recapitulated by spontaneous local and interareal synchronization. Proceedings of the National Academy of Sciences
Fries P (2015) Rhythms for cognition: communication through coherence. Neuron
Lewis CM, Bosman CA, Fries P (2015) Recording of brain activity across spatial scales. Current opinion in neurobiology

References

Lewis, C., Bosman, C.A., Womelsdorf, T., and Fries, P. (2016). Stimulus-induced visual cortical networks are recapitulated by spontaneous local and interareal synchronization. Pnas.

1.4 Phase lags and inter-areal time delays: data and model

Task T3.1.5 Matias Palva (UH), Satu Palva (UH), Viktor Jirsa (AMU)

Overview

Complex phase, amplitude, and cross-frequency correlations of neuronal oscillations are ubiquitous characteristics of emergent brain dynamics. Phase correlations, in particular, are likely to mechanistically underlie the regulation of neuronal communication and may thus be crucial for both functional integration and segregation (Fries, 2015b). The functional consequences of phase correlations are, however, fundamentally determined by the phase lags between the coupled populations. Phase lags are known to depend on axonal conduction delays, coupling strengths, frequencies of the oscillations, and specific patterns of network structures (Klopp et al., 2000; von Stein et al., 2000), and hence it might not be surprising that a large variety of phase-lag observations have been made with pairwise lag assessments by using local field potential or multi-unit activity recordings (Dotson et al., 2014a). While these observations reveal the phenomenological complexity of neuronal phase relations, they leave the large-scale cerebral structure of phase lags undisclosed.

Systematic mapping of phase relationships in large-scale cortical networks is thus an important task but it has remained a difficult one and marred by methodological challenges. The principal challenge is that the commonly used large-scale electrophysiological tools, such as invasive human/primate electrocorticography (ECoG) or non-invasive electro-/magnetoencephalography (EEG/MEG), can yield neither accurate nor well-localized phase estimates because the signals are contaminated by volume conduction and signal mixing (Palva and Palva, 2012). Nevertheless, two new lines of research have advanced solutions to this problem. First, for primates, recording techniques enabling the usage of up to 64 micro-electrodes have been developed to allow concurrent mappings of large cortical regions-of-interest (Dotson et al., 2015). Obviously, this approach is not applicable to humans or easily expandable to primate whole-brain recordings. Second, the development of accurate electrode localization algorithms (Arnulfo et al., 2015a) have opened the possibility to use white-matter referencing (Arnulfo et al., 2015b) in human stereotactical-EEG (SEEG) recordings (Lachaux et al., 2000), which yield undistorted phase estimates unlike the traditional bipolar-referenced SEEG (Arnulfo et al., 2015b). Human SEEG can be recorded from epileptic patients during pre-operative monitoring and yields data on average from 120 electrode contacts in cerebral grey matter.

Large-scale cerebral structure of lags among phase-coupled neuronal oscillations

Satu Palva, Nitin Williams, Gabriele Arnulfo, Spase Petkoski, Viktor Jirsa, J. Matias Palva

Phase lags are poorly charted determinants for the functional impact of phase correlations

Two complementary mechanistic dimensions underlie the processing and representation of information in neuronal circuits. Hierarchical feed-forward routes for sensory feature representations are formed through experience-dependent plasticity and can, via rate coding, signal the presence of well-learned stimuli [1]. Yet, coding schemes based on relatively hardwired structural connectivity have severe capacity limitations and cannot alone explain adaptive perceptual inference and action generation because there is an infinite number of possible constellations of sensory features and appropriate perception-action mappings. As a solution to this problem, the brains appear to use complementary temporal coding mechanisms where the temporal correlations of spikes and phase correlations of oscillating neuronal assemblies signal, *e.g.*, the perceptual relatedness of sensory features and regulate neuronal communication between functionally connected brain regions [2-4], respectively. While the condition-dependent strengths of such synchronization or phase correlations have been extensively studied under various conditions, much less attention has been devoted to understanding the phenomenology and functional significance of the associated time/phase lags even though these lags, *per se*, are crucial for the functional outcome of the coupling. In this review, we provide a summary of prior studies reporting time/phase-lag estimates and an overview of the lines of methodological challenges limiting the acquisition of accurate lag estimates from electrophysiological recordings. We will then address the new vistas on the large-scale organization of phase lags revealed by multi-channel recordings in animals and humans as well as on the insights into the underlying mechanisms given by computational models.

Original observations of near-Zero-lag phase coupling

In many early studies, phase-correlations among neuronal gamma-band (30-90 Hz) oscillations were observed in local field potential (LFP) and multi-unit activity (MUAs) recordings in cat and monkey visual cortices during visual stimulation (for review see [5]). These correlations systematically involved near-zero millisecond time lags. Early neurophysiological and simulation studies suggested that there is a non-linear boost in the postsynaptic impact on target neurons when the presynaptic inputs are synchronized within a few milliseconds [6]. Gamma-band synchronization with time lags << 10 ms thus conceivably endows the oscillating assembly with a competitive advantage in the triggering of action potentials in downstream neurons compared to temporally uncorrelated populations [4, 5, 7]. Substantial experimental evidence supporting this hypothesis has been obtained from multi-unit activity in visual cortex of anaesthetised cat [8]. In the context of vision, gamma-band synchronization and spike-time correlations have been suggested mechanistically to subserve figure-ground separation, sensory feature binding, and bottom-up information representation [9]. In the slower frequencies, such as in the theta-band (4-8 Hz), because of both conduction delays and slower synaptic mechanism, the temporal integration windows were historically thought to be wider and less accurate [10, 11]. It was hence also thought that synchronization in the high gamma frequencies would underlie the integration in small-scale neuronal circuits while that in the slower frequencies could carry out long-range integration [12].

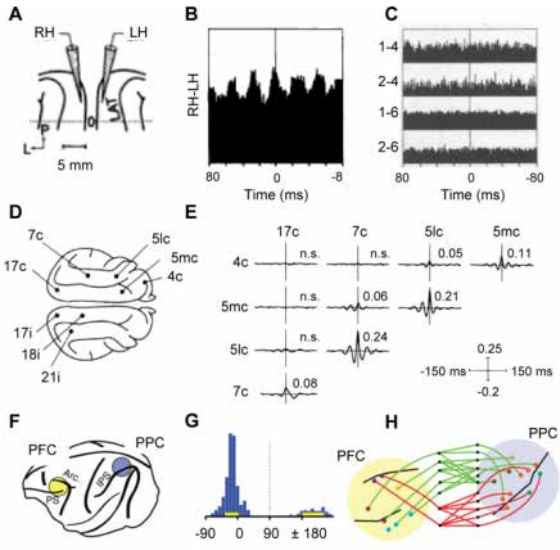


Figure 25: Phase-lags from coupling across spatial scales

(A.) Recording sites from left and right hemispheres of area 17 in anaesthetised cat (left). Cross-correlogram between responses from left and right hemisphere, illustrating near -zero lag phase-coupling in the gamma band. (middle) Cross-correlogram between electrodes in left and right hemisphere respectively, after corpus callosum had been sectioned (right). No phase-coupling is present. Adapted from Engel et al. (1991a) (B.) Recording sites of all locations ipsilateral and contralateral to the paw the cats used to press a lever, in response to visual stimuli (left). Cross-correlogram of areas in left hemisphere of a cat, after lever press. All interactions are zero-lag (right). Adapted from Roelfsema et al. (1997). (C.) Recording sites from brain s of monkeys performing oculomotor delayed match-to-sample task (left). Histogram of phase-angles between PFC (pre-frontal cortex) and PPC (posterior parietal cortex). Note that the distribution of phase-angle is clearly bimodal, with modes around 0 and 180 degrees (middle). Illustration of connections between PPC and PFC, colour-coded by phase-angle (red - ,180 degrees green – 0 degrees) (right). Adapted from Dotson et al. (2014).

Synchronization of neuronal signals is modulated by neuronal oscillations which reflect rhythmic membrane potential fluctuations that are associated with enhanced

neuronal excitability in one phase of the cycle and inhibition in the other [13, 14]. Oscillations thus impose excitability windows that regulate neuronal activity in local networks and thereby both facilitate interactions between areas having an appropriate phase difference and suppress inputs arriving at the inhibitory phase. Zero-phase lag synchronized neuronal oscillations were proposed to have a key mechanistic role also in coordinating inter-areal communication [15]. This "communication through coherence" (CTC) hypothesis suggested that communication is facilitated during the coincident high-excitability phases of neuronal oscillations and likewise suppressed by an anti-phase relationship; coincidence of high- and low-excitability phases.

Zero-lag gamma synchronisation in MUA was observed in cat striate and extra-striate cortical sites separated by 2 to 10 mm [16] [17] (Figure 25A). Coupling lag in these studies was measured with cross-correlation function between MUAs (see Box 2). A number of more recent studies in rodents [18, 19] and primates [20-23] have revealed near zero-phase-lag synchronization among brain oscillations in both attention and memory tasks (Figure 25B).

Accumulating evidence for significant phase lags

In recent years it has become evident that phase-correlations among brain areas at large often involve variable non-zero phase lags in many frequency bands [24-26] (Figure 25C). With this paradigm shift, also some of the small but systematic lags previously considered to be 'near zero' can be seen to build up a picture of a systematic non-zero lag structure of cortical interactions (Figure 26, Supplementary Table). Accordingly, the presence of phase lags has also been included in theoretical frameworks for the functional role of synchronisation [2, 8, 27]. Phase differences arising from conduction delays and emergent dynamics thus play a key role in the functional impact of neuronal phase-coupling. In neuronal oscillations, the excitation inhibition cycle is often skewed and in the gamma frequencies the inhibition is much longer than the excitation period [28]. The revised CTC framework [2] posits that in such a system, phase delays between communicating neuronal groups may enable optimal communication when they match in time with the axonal conduction delays.

Trends in phase lags – Dependence on distance and frequency

In general, it is thought that the phase-lag and corresponding time-delay between two coupled oscillations increases with physical distance between them because of the increases in the neuronal conduction delays. Such correlation between anatomical distance and phase-delay has been indeed observed in gamma-band synchronization. Gamma-band synchronization between V1 and V2 of early visual cortex of macaque monkeys is associated with delays between 0 to 3 ms [29] [30] while that with most electrode pairs less than 10 mm apart is associated with the MUA-LFP and LFP-LFP coupling with up to 5 ms delays [20]. Furthermore, gamma-band synchronization reflecting inter-areal top-down influences is associated with lags up to 6 ms between V4 on V1 of awake cats [31] and delays from 8 to 13 ms in monkey prefrontal-parietal connections [32]. In humans, SEEG has been used to reveal inter-areal synchronization between cortical areas. WM tasks are associated with gamma-band synchronization within the medial temporal lobe [33] and beta-band (15-25 Hz) synchronization within the visual cortex [34]. While medial temporal gamma synchrony was associated with near zero phase lags, visual cortex beta synchronization with electrode separation of 5 cm was associated with variable and clearly non-zero phase-lags. Taken together, time delays appear to increase as a function of anatomical distances in cortical microcircuits.

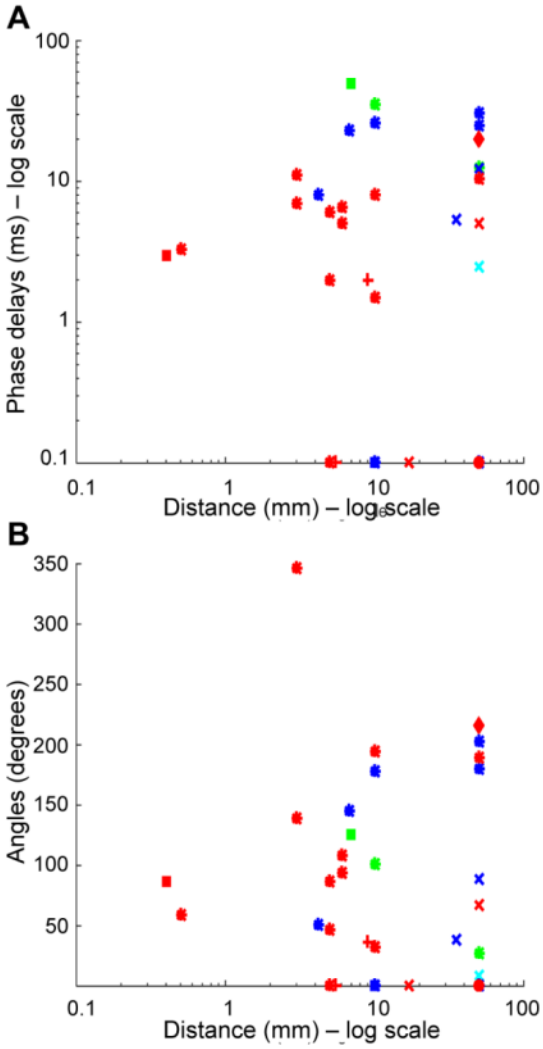


Figure 26: Association of distance vs. phase-lags and phase-angles

(A.) Scatter plot relating distance between recording sites to phase-delays in milliseconds, on log-log scale. Each datum is colour-coded by coupling frequency, i.e. delta (magenta), theta (green), alpha (cyan), beta (blue) and gamma (red). Marker type indicates recording technique used, i.e. multi-unit recordings (+), multi-electrode arrays (*), SEEG (x), ECoG (o), MEG (diamond) and others (square). Most points indicate gamma-band coupling with multi-electrode arrays, at different lags. A few indicating near-zero lag coupling (set to 0.1 ms) is also seen. (B.) Scatter plot relating distance between recording sites to phase-angles in degrees, in semi-log scale. Colour and marker type of data points same as A. Detailed phase-lag and related information on all studies represented available from Table in Supplementary material.

Box 1. Recording technologies to study phase-coupling and phase-lags

EEG – Scalp electrodes record electrical potentials associated with extra-cellular volume currents caused by coherent synaptic inputs to large asymmetric pyramidal neurons mainly in deep cortical layers. Both volume conduction and the distance of the electrodes from the sources introduce linear mixing between the recorded signals, which corrupts both local and inter-areal phase estimates.

MEG – Records weak magnetic fields outside the skull from humans, via superconducting quantum interference devices (SQUIDs). Has a higher spatio-temporal resolution than EEG, but phase-lag estimation with MEG is also vulnerable to signal distortion due to linear mixing effects. Sensitive mainly to sulcal sources.

ECoG – Platinum-iridium or stainless steel electrodes placed sub-durally to record LFP activity from humans. Arrays of grid or strip electrodes typically used. Phase-lag estimation is not as compromised by linear mixing as in EEG, although some mixing from brain tissue still occurs. Less coverage than EEG since sulcal signals are only weakly sampled or missed altogether, and sub-cortical sources are not sampled.

Stereo-tactical EEG – Records LFP activity from humans by insertion of several platinum-iridium shafts (~10) through the cerebral tissue. Phase-lag estimation is not as compromised by linear mixing as in MEG/EEG, although bipolar montages can induce signal distortion. Both cortical and sub-cortical coverage, although coverage is sparser than MEG/EEG, i.e. lower percentage of total LFP activity across the brain is recorded.

Micro-electrode arrays – Dense rectangular arrays of micro-electrodes, typically used on non-human primates to measure both LFP and MUA (and even SUA). Typically confined to a single neuronal volume, advances have enabled coverage of multiple brain areas. Phase-lag estimation is minimally distorted by linear mixing effects. Different laminae can be simultaneously sampled, as can both cortical and sub-cortical sources.

Multi-unit recordings – Record action potentials (specifically, multi-unit activity or MUA) from local clusters of neurons in animals, by inserting conductive micro-electrode (typically platinum or tungsten) near a group of cell membranes. While accurate estimation of phase-coupling and phase-lags is possible, coverage is severely limited..

Single-unit recordings – Records action potentials (specifically, single-unit activity or SUA) from single neurons by manoeuvring the tip of conductive micro-electrode near a single cell membrane. Coverage is even further limited than in multi-unit recordings, although accurate estimation of phase-coupling and phase-lags is possible.

Box 2. Analysis methods for determining phase-lags

Cross-correlation function - Vector of correlations of a signal with time-lagged version of another signal. Lags of zero are included and correlation values calculated up to positive and negative values of a given lag. Coupling delay is taken as the lag at which cross-correlation vector reaches its peak. Typically applied to SUAs and MUAs.

Coherence - Frequency-specific measure of coupling, obtained by normalising cross-spectral density function of two signals by their individual auto-spectral densities. For a given frequency, phase-lag is the angle of the complex-valued cross-spectrum. Typically applied to assess MUA-LFP or LFP-LFP relationships.

Partial Coherence - Frequency-specific measure of coupling between two signals that is independent of the effect of an explanatory signal. Obtaining by subtracting a linear projection of the explanatory signal onto each of the two original signals, and estimating coherence between the residuals. As for coherence, phase-lag is angle of the complex-valued cross-spectrum. Generally applied to LFPs.

Multi-Variate Auto-Regressive (MVAR) modelling - Multi-variate model of brain activity based on Granger causality, relating present value of a given electrode to past values of all electrodes including itself. Maximum coupling delay is equivalent to the estimated model order, usually given by minima of AIC/BIC values over a range of lags. Typically applied to multi-variate LFP data.

Spike-Triggered Averages - Measure of delay between spike and LFPs, obtained by averaging LFPs aligned by time of occurrence of spikes. Coupling delay determined by positive/negative lag corresponding to maxima of LFP average closest to 0. Applied to assess MUA-LFP or SUA-LFP relationships.

Apart from distance, lags between phase synchronized oscillations are influenced by the frequency at which synchronization takes place. In studying coupling between visual cortex and higher cortical areas in awake cats, von Stein et al. (2000) [12] found that coupling at the lower frequencies of theta and alpha bands had 20-50 ms lags, while in the higher gamma band frequencies the phase lags were only 1 □2 ms. Similarly, theta-band synchrony between medial pre-frontal cortex and hippocampus of freely behaving rats has been associated with delays of 50 ms [35]. Compared to monkey

gamma-band synchronization between PFC and PPC [32], beta-band coupling among these areas are associated with much larger delays of 30 ms in awake cats [36]. Time lags thus appear to increase roughly in proportion with the cycle durations indicating and interesting form of scale-freeness that is not trivially explainable by the axonal or synaptic conduction delays.

Anomalous lag findings and the zero-vs-non-zero lag conundrum

Although many studies suggest that phase and time delays are dependent on the anatomical distance and oscillation frequency, there are numerous examples of minimal delays despite a large physical separation or the involvement of slow frequencies. Theta-band (3-9 Hz) coupling between inferior prefrontal cortex and V4 in awake macaque monkeys has been shown to have lags of only 15 ms [37]. In particular, there are a number of studies reporting long-range beta-band synchronization with lags around 0 ms [38]. Zero-phase lag synchrony has been observed between visual and parietal cortices [38] and between somatosensory and parietal areas [39]. Dotson et al. (2014) [40] reported beta-band coupling (8 [illegible] frontal cortex (PFC) and posterior parietal cortex (PPC) of awake macaque monkeys, with bimodal distribution of delays centred around 0 ms and 180 degrees during a working memory WM task (see Figure 25C). 180 degree lags and hence anti-phase synchronization has been observed in PFC-PPC connections also during reaching [41].

However, lags that are much larger than expected by the distance are also frequently observed. For example, gamma band synchronized spikes of a single excitatory and inhibitory neural population are separated by lags of 3.3 ms despite them being only a few hundred microns apart [42]. The relationship between lag and distance also does not seem to hold in gamma-band synchronization between sites up to 900 microns apart but and lags between 7 and 11 ms and hence longer than those in the gamma frequency band [43]. Tallon-Baudry et al. (2004) [44] also report beta-band (15 [illegible] 20 Hz) coupling of sites only 6.7 mm apart in inferior temporal cortex of macaque monkeys performing a working memory task with relatively high coupling delays of around 23 ms.

Phase or time lags between oscillating assemblies cannot thus be 'mechanically' accounted for by considerations of site separation or oscillation frequency. As summarized in Figure 26, the current body of pairwise lag estimate data shows a large diversity of possible phase and time lags. This analysis also shows that it is essentially impossible to draw inferences about the large-scale organization of phase lags from the scattered body of studies examining small numbers of recording sites.

Paradigm shift in mapping the large-scale organization of phase-lags

To reconcile the disparate views on phase and time lags, measuring the phase relationships in large-scale cortical networks is a fundamentally significant task. This endeavor has, however, remained difficult and marred by methodological challenges (see Box 1). There are a number of confounding factors that influence the accuracy of phase- and time-lag estimates.

Bipolar referencing used in both micro-electrode arrays and in SEEG distorts the signal by interference and hence introduce large errors in lags [26]. Micro-electrode arrays also still do not offer whole-brain coverage and can only be used with non-human primates, thus limiting the range of cognitive tasks that can be studied. While ECoG, EEG and MEG provide wider cortical coverage, their spatial resolution is compromised due to linear mixing caused by volume conduction and signal mixing. Linear mixing directly affects both the accuracy of local phase estimates and the estimates of phase differences. Moreover, in ECoG, sulcal signals are confounders or are missed altogether.

However, two recent technical advances: (i) recording LFPs through multi-channel micro-electrode arrays in primates (see Dotson et al. (2015) [24] for review) and (ii) using white-matter referencing for human stereo-tactical EEG (SEEG) recordings [26] have opened novel avenues for accurately determining phase-lags concurrently from many recordings sites in large-scale cortical and subcortical networks.

Dotson and colleagues used micro-electrode arrays to record LFPs from the prefrontal (PFC) and posterior parietal cortex (PPC) of two monkeys during the delay period of a visual working memory (VWM) task. A number of statistically significant connections was then estimated separately for all frequency bands and thereafter the phase-distributions for these connections [24]. Most of the phase correlated between PFC and PPC were found in the theta-, alpha-, and beta-frequency bands. PFC-PPC phase-lags were centred bimodally around zero and 180 degree lags while locally each area was characterized by near-zero lag coupling. These data thus revealed concurrent in-phase and anti-phase coupling but the spread of the phase difference distributions is also suggestive of systematically non-zero lagged phase relations.

Recently, Arnulfo et al. (2015) [26] introduced a new referencing scheme for SEEG where gray-matter contacts are referenced to closest contacts in white matter. This approach both controls volume conduction and minimizes the interference from signals picked up at the reference the white matter does not contain significant bipolar signal generators. In addition, the signals are largely acquired with correct polarity unlike in bipolar where the polarity is essentially random. This is crucial for dissociating true in-phase and anti-phase couplings. In a recent study (Arnulfo, 2016), white-matter referencing was used to systematically map phase and time-delays from resting state activity. The authors found that frequency-dependently, large fractions of cortical phase couplings were associated with a systematic phase difference. Using statistical testing to identify true non-zero lags, the authors found that in low frequencies, most couplings involved significant lags and large phase diversity, but progressively towards the gamma frequencies, zero- and 180-degree phase relationships accounted for up to one half of significant connections. In addition, with increasing frequency, the phase spread of the couplings remained constant, which comprehensively consolidates the prior views of low-frequency oscillations involving unexpectedly large time and phase lags, and gamma oscillations being extremely accurate in their temporal organization.

Analytical and computational studies reveal the structural and mechanistic basis of lags

Experimental studies have demonstrated a diversity of phase relationships among neuronal assemblies, varying with both distance and frequency. Both dependencies have also been observed in a range of computational models.

Network models of coupled oscillators show self-sustained emergent oscillatory dynamics with phase lags near zero or near-180-degree. It is often argued that such behaviour arises from nonlinear couplings, but obviously a more detailed specific understanding on the underlying mechanisms is needed. Analytical and computational investigations suggest that these phase lags are shaped by the interplay of a number of functional properties of the networks, e.g. mean frequencies of the populations, coupling strength. For example, for the network synchronization of two interacting populations of oscillators [45], the phase shift between the mean fields depends on the natural frequencies and the coupling strengths within and between populations. Similar findings were reported for asymmetrically interacting oscillator ensembles [46]. The parameter dispersion of the mean field and mean population frequencies of oscillatory networks equally influences the phase-lags between individual oscillators [47]. Recent modelling studies reveal an influence of topological properties of the networks, which are quantified using graph theoretical metrics such as node strength or node degree. For example, it has been shown that stronger connected network nodes are phase lagging behind the weaker ones [49] suggesting that the relationship between node degree and directionality contributes to the emergent behaviour of network function. Consistent with these findings, Stam et al. showed in a computational model that the phase lead/lag relationship between local node dynamics is correlated with the degree of the node [50]. However, the majority of studies did not systematically address the influence of time delays, which play an increasingly dominant role, together with the structural connectivity, for larger, in particular full brain networks. Usually time delays via signal transmission are assumed to be small. In order to be negligible, they need to be smaller than the half period of the natural oscillators, which is a condition not satisfied for full brain networks, where time delays can reach 200ms.

There is a small number of influential studies investigating how time delays shape network dynamics, its synchronization, and in turn, the phase-lags between nodes. Essentially two limit cases have been considered, either a network of identical oscillators, which are coupled via complex connectivity, or alternatively more complex network nodes, but coupled with simple connectivity such as all-to-all or random couplings. Time delays then enter in both cases as an additional complexification. An early case is a network of identical oscillators coupled with a single delay [51], followed by other studies on heterogeneous networks with single delay [52], and homogeneously distributed delays [53]. These studies all demonstrate that time delays directly change the timing of the interactions between the oscillators and can have highly non-trivial effects on the phase-lags. In particular, for the case of spatially distributed time delays, as relevant for the brain, analytical studies revealed that the delays impose phase-shifted in- and anti-phase clusters of oscillators depending on the spatial distribution. The spatially distributed delays were shown to impose phase-shifted, in- or anti-phase clustering of oscillators depending on the spatial distribution of delays. Anti-phase synchronization has been observed in computational models with different frequency bands [54]. It was found that beside in-phase synchronization within modules, a delay time close to half-period could induce anti-phase synchronization. In full brain network models with realistic connectivity, anti-phase spatio-temporal patterns can emerge from noise-driven transitions between different multi-stable cluster synchronization states, with a two-community network structure [56].

Functional Implications

Several lines of experimental and theoretical evidence thus converge to show that both zero and non-zero lag phase coupling are pervasive in neuronal systems at all scales from millimetres to fronto-posterior connections. Neuronal communication may thus be dynamically regulated by two distinct mechanisms. In the first, near-zero lag synchronization of presynaptic neuronal assemblies endows them to have a greater impact on the postsynaptic target neuron than asynchronous assemblies [9, 11, 57]. In the second, phase correlation of pre- and post-synaptic neurons facilitates or blocks the communication depending on the excitability phase of the post-synaptic neuron at the time of arrival of the presynaptic inputs and hence, depending on the phase difference and conduction delay between these neurons [2]. Both mechanisms support adaptive and rapidly reconfigurable temporal coding and appear likely to co-operate. New empirical evidence on the large-scale structure of phase lags in cortical and subcortical networks sheds light on the conditions wherein either or both of these mechanisms may be at play.

As shown by several studies here, zero-lag phase interactions characterize oscillatory synchronization in the gamma but to a lesser extent also in other frequency bands. Furthermore, although zero-lag synchronization is more prevalent across short than large anatomical distances, it has been also observed also between widely separated cortical structures such as the prefrontal and posterior parietal cortices. Hence, zero-lag synchronization is a common characteristic of systems-level neuronal activity and cannot be attributed to single functions, frequencies, or brain systems. Interestingly, as several lines of evidence also point to systematic anti-phase synchronization, there is a solid experimental basis for the phase correlations to support both neuronal integration, by zero-lag coupling, and segregation, by anti-phase coupling [58].

References

1 Singer, W. (1995) Development and plasticity of cortical processing architectures. *Science* 270, 758-764

2 Fries, P. (2015) Rhythms for Cognition: Communication through Coherence. *Neuron* 88, 220-235

3 Singer, W. (2013) Cortical dynamics revisited. *Trends Cogn Sci* 17, 616-626

4 Uhlhaas, P.J. *et al*. (2009) Neural synchrony in cortical networks: history, concept and current status. *Front Integr Neurosci* 3, 17

5 Singer, W. (1999) Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24, 49-65, 111-25

6 Softky, W.R. and Koch, C. (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J Neurosci* 13, 334-350

7 Freiwald, W.A. *et al*. (1995) Stimulus dependent intercolumnar synchronization of single unit responses in cat area 17. *Neuroreport* 6, 2348-2352

8 Nikolic, D. *et al*. (2013) Gamma oscillations: precise temporal coordination without a metronome. *Trends Cogn Sci* 17, 54-55

9 Singer, W. and Gray, C.M. (1995) Visual Feature Integration and the Temporal Correlation Hypothesis. *Annu Rev Neurosci* 18, 555-586

10 Konig, P. *et al*. (1995) How Precise is Neuronal Synchronization. *Neural Comput* 7, 469-485

11 Konig, P. *et al*. (1996) Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends Neurosci* 19, 130-137

12 von Stein, A. *et al*. (2000) Top-down processing mediated by interareal synchronization. *Proc Natl Acad Sci U S A* 97, 14748-14753

13 Pastor, M.A. *et al*. (2002) Activation of human cerebral and cerebellar cortex by auditory stimulation at 40 Hz. *J Neurosci* 22, 10501-10506
14 Schroeder, C.E. and Lakatos, P. (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci* 32, 9-18
15 Fries, P. (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci* 9, 474-480
16 Engel, A.K. *et al*. (1991) Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proc Natl Acad Sci U S A* 88, 6048-6052
17 Engel, A.K. *et al*. (1991) Interhemispheric Synchronization of Oscillatory Neuronal Responses in Cat Visual-Cortex. *Science* 252, 1177-1179
18 Colgin, L.L. *et al*. (2009) Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature* 462, 353-357
19 Diba, K. *et al*. (2014) Millisecond timescale synchrony among hippocampal neurons. *J Neurosci* 34, 14984-14994
20 Womelsdorf, T. *et al*. (2007) Modulation of neuronal interactions through neuronal synchronization. *Science (New York* 316, 1609-1612
21 Womelsdorf, T. *et al*. (2006) Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature* 439, 733-736
22 Salazar, R.F. *et al*. (2012) Content-specific fronto-parietal synchronization during visual working memory. *Science* 338, 1097-1100
23 Buschman, T.J. and Miller, E.K. (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860-1862
24 Dotson, N.M. *et al*. (2015) Methods, caveats and the future of large-scale microelectrode recordings in the non-human primate. *Front Syst Neurosci* 9, 10.3389/fnsys.2015.00149
25 Maris, E. *et al*. (2016) Diverse Phase Relations among Neuronal Rhythms and Their Potential Function. *Trends Neurosci*
26 Arnulfo, G. *et al*. (2015) Phase and amplitude correlations in resting-state activity in human stereotactical EEG recordings. *Neuroimage* 112, 114-127
27 Steinmetz, P.N. *et al*. (2000) Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* 404, 187-190
28 Buzsaki, G. and Wang, X.J. (2012) Mechanisms of gamma oscillations. *Annu Rev Neurosci* 35, 203-225
29 Jia, X. *et al*. (2013) Gamma and the Coordination of Spiking Activity in Early Visual Cortex. *Neuron* 77, 762-774
30 Besserve, M. *et al*. (2015) Shifts of Gamma Phase across Primary Visual Cortical Sites Reflect Dynamic Stimulus-Modulated Information Transfer. *PLoS Biol* 13, e1002257
31 Salazar, R.F. *et al*. (2004) Directed interactions between visual areas and their role in processing image structure and expectancy. *Eur J Neurosci* 20, 1391-1401
32 Gregoriou, G.G. *et al*. (2009) High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science (New York* 324, 1207-1210
33 Fell, J. *et al*. (2001) Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. *Nat Neurosci* 4, 1259-1264
34 Tallon-Baudry, C. *et al*. (2001) Oscillatory synchrony between human extrastriate areas during visual short-term memory maintenance. *J Neurosci* 21, art. no.-RC177
35 Siapas, A.G. *et al*. (2005) Prefrontal phase locking to hippocampal theta oscillations. *Neuron* 46, 141-151
36 Pesaran, B. *et al*. (2008) Free choice activates a decision circuit between frontal and parietal cortex. *Nature* 453, 406-409
37 Liebe, S. *et al*. (2012) Theta coupling between V4 and prefrontal cortex predicts visual short-term memory performance. *Nat Neurosci* 15, 456-62, S1-2
38 Roelfsema, P.R. *et al*. (1997) Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature* 385, 157-161
39 Witham, C.L. *et al*. (2007) Cells in somatosensory areas show synchrony with beta oscillations in monkey motor cortex. *Eur J Neurosci* 26, 2677-2686
40 Dotson, N.M. *et al*. (2014) Frontoparietal correlation dynamics reveal interplay between integration and segregation during visual working memory. *J Neurosci* 34, 13600-13613
41 Stetson, C. and Andersen, R.A. (2014) The parietal reach region selectively anti-synchronizes with dorsal premotor cortex during planning. *J Neurosci* 34, 11948-11958
42 Vinck, M. *et al*. (2013) Attentional modulation of cell-class-specific gamma-band synchronization in awake monkey area v4. *Neuron* 80, 1077-1089
43 Maris, E. *et al*. (2013) Rhythmic neuronal synchronization in visual cortex entails spatial phase relation diversity that is modulated by stimulation and attention. *Neuroimage* 74, 99-116
44 Tallon-Baudry, C. *et al*. (2004) Oscillatory synchrony in the monkey temporal lobe correlates with performance in a visual short-term memory task. *Cereb Cortex* 14, 713-720
45 Montbri\'o Ernest *et al*. (2004) Synchronization of two interacting populations of oscillators. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 70, 1-4
46 Sheeba, J.H. *et al*. (2008) Routes to synchrony between asymmetrically interacting oscillator ensembles. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 78, 2-5
47 Petkoski, S. *et al*. (2013) Mean-field and mean-ensemble frequencies of a system of coupled oscillators. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 87, 1-12
48 Hong, H. and Strogatz, S.H. (2011) Kuramoto model of coupled oscillators with positive and negative coupling parameters: An example of conformist and contrarian oscillators. *Phys Rev Lett* 106, 1-4
49 Moon, J.Y. *et al*. (2015) General relationship of global topology, local dynamics, and directionality in large-scale brain networks. *PLoS Computational Biology* 11, 1-21
50 Stam, C.J. and Straaten, E.C.W.v. (2012) Go with the flow: Use of a directed phase lag index (dPLI) to characterize patterns of phase relations in a large-scale model of brain dynamics. *Neuroimage* 62, 1415-1428
51 Yeung, M. and Strogatz, S. (1999) Time Delay in the Kuramoto Model of Coupled Oscillators. *Phys Rev Lett* 82, 648-651
52 Choi, M. *et al*. (2000) Synchronization in a system of globally coupled oscillators with time delay. *Physical Review.E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 61, 371-381
53 Lee, W.S. *et al*. (2009) Large coupled oscillator systems with heterogeneous interaction delays. *Phys Rev Lett* 103, 4-7
54 Li, D. and Zhou, C. (2011) Organization of Anti-Phase Synchronization Pattern in Neural Networks: What are the Key Factors? *Frontiers in Systems Neuroscience* 5, 100
55 Deco, G. *et al*. (2009) Key role of coupling, delay, and noise in resting brain fluctuations. *Proc Natl Acad Sci U S A* 106, 10302-10307
56 Honey, C.J. *et al*. (2007) Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc Natl Acad Sci U S A* 104, 10240-10245
57 Azouz, R. and Gray, C.M. (2003) Adaptive coincidence detection and dynamic gain control in visual cortical neurons in vivo. *Neuron* 37, 513-523
58 Deco, G. *et al*. (2015) Rethinking segregation and integration: contributions of whole-brain modelling. *Nat Rev Neurosci* 16, 430-439

Glossary

Bipolar montages: Recording arrangement whereby each electrode contact in a linear shaft is referenced to the one neighbour contact.

Emergent dynamics: Dynamics arising from the interaction of lower-level constituents, but not reducible to them.

Graph-theoretic property: Feature of an abstraction of a complex system into a set of nodes and edges. For example, the average number of edges per node, i.e. the mean degree.

Mean-field: Single state-variable describing for e.g. the dynamics of a neuronal population, by approximating the effect of interactions between its many constituent elements.

Generally employed in computational modelling, to gain insight into system behaviour at relatively low computational cost.

Neuronal oscillations: Periodic signal, typically observed at the level of neuronal populations. Oscillations are generated by the interaction between excitatory pyramidal neurons and inhibitory interneurons.

Phase-correlation: Systematic dependency between the instantaneous phases of two oscillators. Also referred to as phase-synchronisation or phase-coupling.

Phase-lag: The angle by which an oscillator lags or leads another, when their instantaneous phases do have a systematic dependency.

Volume conduction: Transmission of electric or magnetic fields from a primary current source to measurement sensors, via intermediate tissue such as the brain, skull and scalp.

Supplementary Table 1. Tabulation of phase-lag findings from neuro-physiological data

	Phase lags	**Freq. bands**	**Distance**	**Recording**	**Analysis method**	**Species**	**Brain region**	**Task (Animal state)**
Gray et al. (1989)	~0 ms	40-60 Hz	2-7 mm	Multi-unit recordings (MUAs)	Cross-correlation function	Cat	Visual cortex (area 17)	Visual stimulation with moving light bars (Anaesthetised)
Engel et al. (1991)	around 2 ms	40-60 Hz	8-10 mm	Multi-unit recordings (MUAs)	Cross-correlation function	Cat	Area 17 and Posteromedial lateral suprasylvian (PMLS)	Visual stimulation with moving light bars (Anaesthetised)
Engel et al. (1991a)	0 ms	40-60 Hz	~5 mm	Multi-unit recordings (MUAs)	Cross-correlation function	Cat	Area 17 of left & right hemispheres	Visual stimulation with moving light bars (Anaesthetised)
Livingstone et al. (1996)	3 ms	70-90 Hz	300-500 microns	Enamel-coated tungsten electrodes (LFPs and MUAs)	Cross-correlation	Squirrel monkey	electrodes were in deep layer 4B and superficial layer 2/3 (early visual cortex)	Visual stimulation (Anaesthetised)
Fries et al. (2008)	~0 ms	30-70 Hz	4-6 mm	Multi-electrode arrays (LFPs and MUAs)	Coherence (b/n MUAs)	Macaque monkey	V4	Visual stimulation, Selective visual attention (Awake)
Roelfsema et al. (1997)	0 ms	20-25 Hz	>1 cm	Trans-cortical electrodes (LFPs and MUAs)	Cross-correlation function	Cat	Parietal & Motor cortices (areas 7 and 5l)	Visuo-motor integration task (Awake)
Jia et al. (2012)	3 ms (between spikes), 5-8 ms (between LFPs)	30-50 Hz	~6 mm	Multi-electrode arrays in V1 and tetrodes in V2 (LFPs, MUAs and SUAs)	Cross-correlation function (b/n spikes), Coherence (b/n LFPs)	Macaque monkey	V1 and V2 (early visual cortex)	Visual stimulation (Anaesthetised)

Grothe et al. (2012)	~8 ms	45-90 Hz	~10 mm	Micro-electrode array (LFPs, MUAs and SUAs)	Spike-triggered average	Macaque monkey	V1 and V4 (visual cortex)	Selective visual attention task (Awake)
Gregoriou et al. (2009)	8-13 ms	40-60 Hz (but also theta and beta frequencies)	>5 cm	Multi-electrode arrays (LFPs and MUAs)	Spike-field coherence	Monkey	Frontal eye field (in PFC) and V4	(Attention task) Awake
Brovelli et al. (2004)	up to 26 ms	16-22 Hz	>1 cm	Micro-electrode array (LFPs)	Coherence	Macaque monkeys	Primary somatosensory cortex, Motor cortex, Inferior Posterior Parietal Cortex	Motor maintenance (i.e. press hand lever) (Awake)
Siapas et al. (2005)	50 ms	4-10 Hz	7 mm	Tetrodes (LFPs, SUAs)	Spike-field coherence	Rat	Pre-frontal cortex, Hippocampus	Freely behaving (Awake)
Witham et al. (2007)	0 ms	~17.5 Hz	>1 cm	Micro-electrode arrays (LFPs, SUAs)	Coherence, Phase Slope Index	Macaque monkeys	Somatosensory cortex (spikes), Parietal cortex (spikes) ,Motor cortex (LFPs)	Finger-movement task (Awake)
Womelsdorf et al. (2007)	5 ms	40-80 Hz	~2 to 10 mm	Micro-elecrode array (LFPs, MUAs)	Coherence	Monkey, Cat	Cat - area 17, area 18 and area 21a Monkey - V1, V4	Visual stimulation with moving gratings (Awake)
Bosman et al. (2012)	>0 ms	60-80 Hz	>1 cm	ECoG grid	Coherence, Granger causality	Monkey	V1 and V4 of visual cortex	Attentional stimulus selection (Awake)
Baldauf & Desimone (2014)	20 ms	60-100 Hz	>5 cm	MEG (LFPs)	Coherence, Phase Slope Index	Humans	IFJ, FFA and PPA	Object-based attention, i.e. faces and houses (Awake)
von Stein et al. (2000)	20-50 ms for theta/alpha and 1-2 ms for gamma	4-12 Hz, 20-100 Hz	7-10 mm (for area 7-area 5) and > 1 cm (for area 17-area 7, area 17-area 5)	Micro-electrode arrays (LFPs)	Cross-correlation function	Cats	area 17 (primary visual cortex), area 7 (visual and polysensory responses) and area 5 (sensory and sensory-motor responses)	GO/NO-GO task of stimulus perception (Awake)
Tallon-Baudry et al. (2001)	5.4 ms, 12.4 ms	15-25 Hz	3.5 cm, 5 cm	SEEG (LFPs)	Phase Locking Value	Humans	Extrastriate visual areas	Visual short-term memory task (Awake)
Tallon-Baudry et al. (2004)	8 ms, 23 ms	15-20 Hz	4.2 mm, 6.7 mm	Micro-electrode arrays	Phase Locking Value	Macaque monkeys	posterior infero-temporal cortex (near	Visual short-term memory task (Awake)

				(LFPs)			superior temporal sulcus – STS)	
Bastos et al. (2015b)	>0 ms	~ 4 Hz and ~ 60-80 Hz feed-forward influences, 14-18 Hz feedback influences	>1 cm	ECoG grid (LFPs)	Conditional Granger causality	Macaque monkeys	Visual cortex (8 areas – V1, V2, V4, TEO, DP, 7A, 8L, 8M)	Visuo-spatial attention task (Awake)
Vinck et al. (2013)	3.3 ms	30-70 Hz	100s of microns	Micro-electrode array (LFPs, MUAs and SUAs)	Phase Slope Index	Macaque monkeys	Visual cortex (area V4, between excitatory and inhibitory cells)	Selective visual attention (Awake)
Maris et al. (2013)	7-11 ms	55 Hz, 87.5 Hz	up to ~3 mm	Micro electrode arrays (LFPs and MUAs)	Coherence	Macaque monkeys	Visual cortex (area V4)	Resting-state, visual stimulation, attention (Awake)
Dotson et al. (2014)	0 ms, ~31.2 ms	8-25 Hz	< 1 cm (within PPC), >5 cm (between PFC and PPC)	Micro-electrode array (LFPs)	Cross-correlation function	Macaque monkeys	Pre-frontal cortex (PFC), Posterior parietal cortex (PPC)	Oculo-motor, delayed match-to-sample task (Awake)
Stetson et al. (2014)	~25 ms	15-30 Hz	>5 cm	Micro-electrode array (LFPs, SUAs)	Partial coherence	Macaque monkeys	Parietal reach region (PRR) and Dorsal Premotor cortex (PMd)	Movement planning (Awake)
Liebe et al. (2012)	10-15 ms	3-9 Hz	>5 cm	Micro-electrode array (LFPs, SUAs)	Phase Locking Value (PLV)	Macaque monkeys	IPF (Inferior Prefrontal Cortex) and V4 in visual cortex	Visual short-term memory task (Awake)
Salazar et al. (2004)	up to 6 ms	20-60 Hz	5 mm	Micro-electrode array (LFPs)	Multi-Variate Auto-Regressive (MVAR) modelling	Cats	Primary (A17/18) and higher visual areas (A21)	Movie and pink noise visual stimulation (Awake)
Besserve et al. (2015)	~2 ms	50-80 Hz	5 mm	Micro-electrode array (LFPs, MUAs)	Phase Locking Value	Macaque monkeys	V1	Movie visual stimulation (Anaesthetised)
Aoki et al. (1999)	~0 ms	30-40 Hz	>5 cm	ECoG grid	Cycle-triggered averages, Cross-spectra	Humans	Forearm sensori-motor cortex	3 visuomotor tasks, i.e. target tracking, finger threading and finger sequencing (Awake)
Klopp et al.	1-4 ms	8-12 Hz,	1 to 9	SEEG	Coherence	Humans	Fusiform gyrus to	Memory task for faces

(2000)	(gamma), 1-9 ms (alpha)	30-45 Hz	cm				several other sites including: anterior cingulated gyrus, posterior cingulate gyrus, anterior and posterior hippocampus, lingual gyrus, supre-marginal gyrus etc.	(Awake)
Fell et al. (2001)	~0 ms on average	32-48 Hz	0.8 to 2.6 cm	SEEG	Phase Locking Value	Humans	Rhinal cortex, Hippocampus	Word memorisation task (Awake)

Data set: Map of human inter-areal connectivity and phase lags based on resting-state SEEG

In this project, we have assessed the large-scale phase-lag structure of human cortical and subcortical network oscillations by using a large set of resting-state SEEG data. These data open a novel and comprehensive view into the organizational principles of neuronal network oscillations.

Objective

This report is the human SEEG-based experimental part of the Deliverable for our objective, followed by the theoretical section based upon modelling of network oscillations within a large-scale connectivity framework:

> *The objective of this work package is to understand the nature of spontaneous brain activity in biologically realistic brain network models with concurrent oscillations in multiple frequency bands. To this end, we exploit the first comprehensive estimates of the true time and phase lags between interacting human cortical regions.*

Design and Methods

Subjects, data acquisition, and data provenance. We collected data from 67 epileptic patients undergoing evaluation prior a surgical ablation of the epileptic zone (EZ). These subjects were selected from among 76 consecutive patients in Niguarda hospital, Milan, Italy (by Dr. L. Nobili and A. Rubino) of which 9 were rejected because of significant cortical lesions or malformations. The data were pre-processed and analysed in the Neuroscience Center, University of Helsinki, Finland (by Dr. G. Arnulfo, J. Hirvonen, S. Wang, Dr. A. Zhigalov, Dr. H. Eyherabide, and Dr. N. Williams). The data were acquired in a project funded by the Academy of Finland (J.M. Palva and S. Palva).

10-minute long resting-state SEEG recordings with eyes closed but without sleeping were acquired with a 192-channel SEEG amplifier system (NIHON-KOHDEN NEUROFAX-110) at a sampling rate of 1 kHz. The subjects were uninterrupted during the recordings and there were no active distractions (television, etc.).

Data characterization and secondary aims. Using the first 22 subjects of the present cohort, we have conducted proof-of-concept and data characterization analyses in order to assess quantitatively the inter-areal phase and amplitude correlations (Arnulfo et al., 2015b), the dynamic states (Zhigalov et al., 2015), and their mutual relationship (Zhigalov et al., PLoS Biol, *revision submitted*). This project also aimed to assess the phase correlation and lag networks underlying threshold-stimulus detection task (TSDT) execution. MEG data for somatosensory (Hirvonen and Palva, 2015) and visual TSDTs (Hirvonen et al., 2016, *in preparation*, respectively) are recorded and the local and inter-areal phase coupling correlates of conscious perception mapped. The SEEG counterpart of this experiment has been delayed and the data acquisition is still ongoing because of delays in the handling of the ethical application at Niguarda, lower than expected patient throughput, and temporary cessation of all SEEG recordings during 2015 caused by the breaking of the SEEG electrode implantation robot.

Preprocessing. For each subject, the SEEG electrode contacts were localized with an automatic segmentation algorithm from CT images and co-localized with anatomical MRI

(Arnulfo et al., 2015a). The anatomical MRIs were segmented to obtain surface models of the cortical white-grey matter and pial surfaces with the Freesurfer software. Cortical surfaces were then parcellated with both Desikan-Killiany and Destrieux atlases while for the labeling of subcortical structures, the volumetric MRI segmentation is used (Desikan et al., 2006a; Destrieux et al., 2010).

Using this localization, we identified contacts in cortical and subcortical grey matter and used the closest contacts in underlying white matter for re-referencing the original 'monopolar' SEEG time series. The electrode contact time series analyzed here thus reflected signals from local grey matter sources with minimized volume conducted components and no introduction of confounding grey matter signals through referencing (Arnulfo et al., 2015b).

Analyses of phase lags at single subject level. All analyses of phase correlations and lags were performed between the raw grey-matter electrode contact time series from individual subjects. The time series were FIR filtered and transformed to complex time series with the Hilbert transform. We first evaluated 'static' functional connectomes by obtaining pairwise complex phase-locking values (cPLV) across all contact pairs for the whole 10 min time series and all frequencies. Contact pairs with shared reference contacts were excluded from all analyses. From the static connectomes, we chose for a time-windowed analysis contact pairs exhibiting significant phase coupling ($p < 0.05$, no multiple comparisons correction, estimated against time-rotated surrogate data to maintain auto-correlation-caused redundancies).

While the static cPLV connectome yields estimates of the overall phase lags, it is a composite of possibly many distinct and dynamically variable phase lags and is hence *a priori* uninformative. We thus assessed the phase-lags with a time-windowed analysis so that phase-coupling between contact pairs was measured in windows of five oscillation cycles with cPLV and windows with PLV > 0.85 were ultimately selected for the analyses of phase lags. A high PLV threshold ensured that the chosen windows reflected strong neuronal phase coupling with a consistent phase relationship. For all contact pairs yielding more than 30 significant windows, the phases of these cPLVs were the pooled across and constituted the contact-pair-specific phase-difference distribution (PDD).

We assessed three statistical properties of these PDDs. First, "systematicity" was measured with the Rayleigh test and indicated whether the phase lags from different time windows were systematically centered on a given lag or distributed uniformly. Second, "non-zero-lagness" was measured by comparing the mean phase across all time windows against the distribution of such mean phases obtained with identical procedures from surrogate data with matched coupling at zero phase lag. Third, the PDD width was compared against the corresponding widths of surrogate data to estimate whether the measured PDDs were a likely to be composites of multiple interactions or explainable by a single one.

Analyses of phase lags at group level. For group analyses, we pooled the means of time-window phase estimates of significantly systematic contact pairs from all subjects into a parcel-parcel connectomes for the Desikan-Killiany (Desikan et al., 2006a) and Destrieux atlases (Destrieux et al., 2010). For each parcel-parcel pair, we then estimated with the Rayleigh test whether the phase lags were systematic or uniformly/randomly distributed. A significant observations here thus indicates that at the group level, a given parcel-parcel pair exhibits a systematic phase lag.

Results

Sampling statistics and static connectivity. In total, the cohort yielded 5710 grey matter electrode contacts of which 5282 were in neocortical and 428 in subcortical or

hippocampal structures. This sampling gave a total of 337946 contact pairs (285439 among cortical, 3937 among subcortical, and 48570 between these two). Collapsing these data to the Desikan-Killiany parcellation showed that most cortical regions were extensively sampled (Figure 27A) and that more than 57 % of parcel pairs were sampled by at least 10 contact pairs (Figure 27B). The least sampled parcel pairs were almost exclusively interhemispheric (Figure 27C), because SEEG electrode contacts are largely implanted unilaterally and hence undersample bilateral connections.

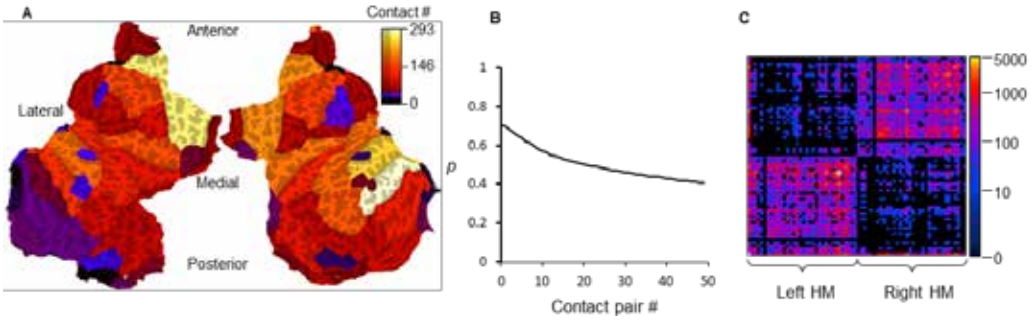


Figure 27: Sampling statistics

A) Inflated cortical surfaces show the number of electrodes (color scale) in the parcels of the Desikan-Killiany atlas. B) The fraction of parcel pairs (y-axis) sampled by N or more electrode pairs (x-axis). C) The parcel-parcel sampling matrix shows the number of electrode pairs (color scale) for pair of parcels.

We first assessed the strength and extent of static, i.e., time-averaged, phase correlations by pairwise PLV estimates (Lachaux et al., 1999) between all electrode contacts. We divided the contact pairs into three groups based on their Euclidian distance: short (0-34 mm), medium (34-53 mm), and long (53-137 mm). Corroborating our prior findings, the strength of phase correlations, as indicated by the mean PLV, decreases systematically with distance as well as with frequency for frequencies above ~10 Hz (Figure 28A). The extent of phase correlations was indexed by the fraction of couplings exceeding a surrogate-data-derived significance threshold corresponding to $p < 0.01$ (uncorrected). These fractions were considerably large (Figure 28B), which together with the small mean PLVs indicates that the brains operate in regime of widespread functional coupling but low overall coherence.

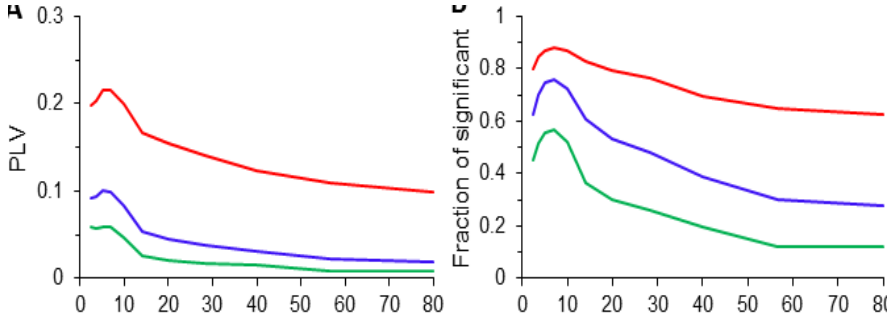


Figure 28: Strength and extent of static phase correlations

A) Mean PLV for short (black-), medium- (red), and long-(green) range electrode pairs. B) Fraction of electrode pairs exhibiting significant ($p < 0.01$) phase correlations.

Dynamic phase-lag analyses. For each frequency band, all electrode pairs exhibiting significant coupling were subjected to a time-window-based analysis that aimed to identify brief five-cycle-long moments of strong phase coupling. Selecting then the electrode pairs where an adequate number (> 30) of significant time windows were found, we first asked whether the phase lags observed in these windows (Figure 29A) would be systematic or

uniformly distributed. Surprisingly, we found that essentially all low- (2.5-12 Hz) and a majority of the high- (12-80 Hz) frequency couplings indeed exhibited a specific systematic phase lag (Figure 29B) indicating that even in the supposedly highly variable resting state, emergent brain dynamics are characterized by an organized mosaic of inter-areal phase relationships.

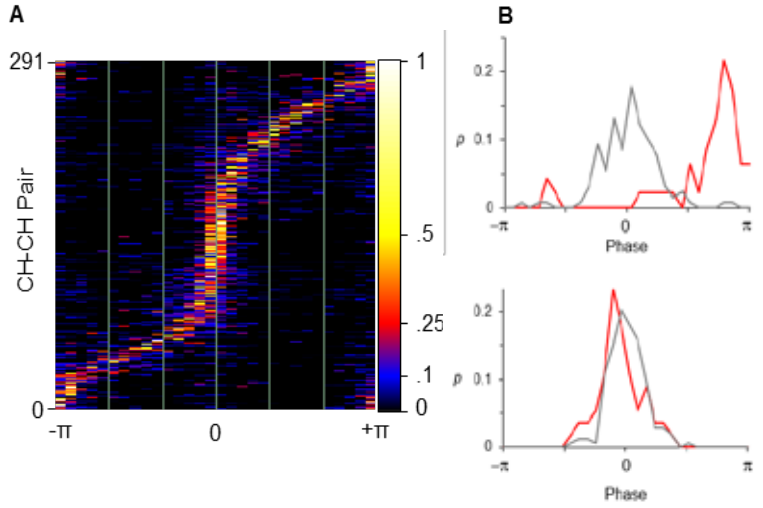


Figure 29: Systematic phase lags in the human brain

A) Phase lag distributions (color scale) of all significant electrode pairs (y-axis) of a representative subject for 10 Hz oscillations. B) Example non-zero (*upper panel*) and zero- (*lower panel*) lag distributions of two electrode pairs (red) and their surrogates (gray) from the data in panel A.

The individual phase lag distributions suggested that in a subset of electrode pairs, the phase lags would be very near zero or 180 degrees. We used zero-lag correlated and coupling-strength matched surrogate data to test whether the lags in these and other pairs were explainable as originating from a true zero-lag phase distribution (Figure 30A). We found that depending on frequency but essentially independently of distance, 50-70 % of the systematic couplings had a lag significantly deviating from zero (Figure 30).

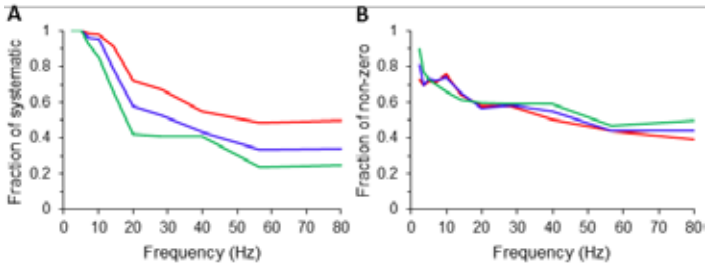


Figure 30: Systematic zero- and non-zero-lagged phase couplings are salient between electrode pairs

A) Fraction of electrode pairs in the complete cohort exhibiting a significant ($p < 0.05$, uncorrected) systematic phase lag. B) Fraction of systematic electrode pairs exhibiting significantly ($p < 0.05$, uncorrected) non-zero phase lag.

Systematic phase lags in cortical parcellations. The lead results so far have shown that phase correlations in human brains are largely systematic and characterized by both near-zero and significantly non-zero phase lags. A key question is whether these phase

relationships are consistent across subjects. We obtained the mean phase across time-window phase estimates for each systematic electrode pair and pooled these phase estimates to become parcel-parcel phase estimates. To ensure adequate statistics, we required the parcel pairs to be sampled by a minimum of ten electrode pairs. For the Desikan-Killiany parcellation, with only 68 parcels, surprisingly large fractions of parcel pairs, 33-56 %, exhibited significantly systematic phase lags across the pooled electrode pairs (Rayleigh test, $p < 0.05$, uncorrected; Figure 31A). Visualizing the parcel-pair phase lag distributions showed that both clearly zero-lagged and systematically non-zero lagged parcel pairs were robustly observable (Figure 31B).

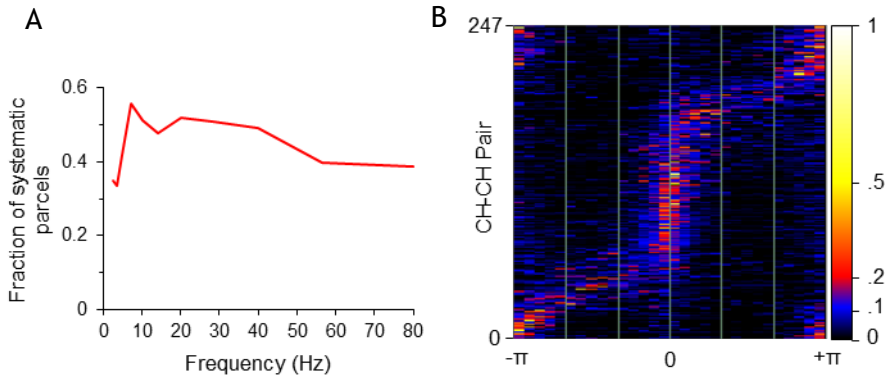


Figure 31: Zero- and non-zero-lagged phase couplings are systematic between large cortical parcels.

A) Fraction of Desikan-Killiany parcel-pairs exhibiting a systematic phase lag across the contributing electrode pairs. B) Phase lag distributions (color scale) of all significantly systematic parcel pairs for 10 Hz oscillations.

Data Provenance

The raw data have been acquired in a collaboration project comprising Matias PALVA, UH, and Lino NOBILI, Claudio Munari Epilepsy Surgery Centre, Niguarda Hospital, Italy.

Delivery and data release

The adjacency and all supporting matrices for phase correlation strength and phase estimates produced in this project are downloadable from the URL: http://sp3.s3.data.kit.edu/3.1.5/ and are described in a Dataset Information Card: Human Connectome of Phase Lags.

Discussion

We have analysed an unprecedentedly large set of SEEG data in order to systematically map, for the first time, cerebral inter-areal phase relationships. This was, in part, also enabled by the usage of a novel white-matter referencing scheme that yielded volume-conduction controlled local field potential recordings from known depths of grey matter without referencing-caused temporal distortions. As summarized in our review, a large body of literature based on pairwise phase estimates shows that phase couplings among neuronal assemblies may involve either near-zero and -180-degree (Dotson et al., 2014b; Roelfsema et al., 1997), or clearly in between, non-zero phase lags (Baldauf and Desimone, 2014; Tallon-Baudry et al., 2001; Womelsdorf et al., 2007). Whether ongoing brain activity at large is dominated by one or another kind of coupling is an important constraint both for conceptual models of the functional significance of phase correlations

and for computational models of large-scale dynamics. To this end, we used a novel statistical approach for dissociating significantly non-zero lagged couplings from those that could not be discerned from an equally sampled true zero-lagged coupling. We found that both zero- and non-zero-lag couplings were saliently observable both in individual electrode pairs and between cortical parcels at the group level. Interestingly, low-frequency oscillations exhibited greater variability in phase lags and greater fractions of non-zero lag couplings than high-frequency oscillations. Overall, this dataset will be a treasure trove for understanding the high-resolution architecture of inter-areal interactions and the time/phase lags therein.

Self-analysis of the value and completeness of data. We achieved our primary objective: a comprehensive description of cortical and subcortical inter-areal phase relationships in the human brain. Outside of our HBP project, both resting- and task-state data acquisition and analyses are still ongoing, and will be integrated with the connectome released in the present study. This will improve the anatomical resolution and inferential value of the phase connectome. With the resolutions used here, however, the coverage given by the 337946 acquired grey-matter electrode contact pairs is more than adequate and hence the data can be considered complete. The value of these data is, in our humble opinion, immense. Human electrophysiological resting-state data of this quality, anatomical accuracy, and quantity has never been acquired – typical publications in the field use 5-10 times smaller subject cohorts, poor anatomical accuracy, and no shareable anatomical descriptions. The connectomes that we are publishing are represented in an anatomical framework directly usable in any human neuroscience laboratory. These data yield the very first large-scale (whole-brain scale) mapping of phase-accurate LFPs from any mammal – the value for understanding the large-scale brain dynamics is hence great.

Usage of data and collaboration. So far these data have been used by the consortium partners. The connectomes and associated metadata will be released to become freely usable when the findings have been accepted for publication. These data also support an ongoing collaboration with SP4, Gustavo Deco.

List of publications.

Arnulfo, G., Hirvonen, J., Nobili, L., Palva, S., & Palva, J. M. (2015). Phase and amplitude correlations in resting-state activity in human stereotactical EEG recordings. NeuroImage, 112, 114–127. http://doi.org/10.1016/j.neuroimage.2015.02.031

Arnulfo, G., Narizzano, M., Cardinale, F., Fato, M. M., & Palva, J. M. (2015). Automatic segmentation of deep intracerebral electrodes in computed tomography scans. BMC Bioinformatics, 16(1), 1–12. http://doi.org/10.1186/s12859-015-0511-6

Zhigalov A, Arnulfo G, Nobili L, Palva S, Palva JM (2015) Relationship of fast- and slow-timescale neuronal dynamics in human MEG and SEEG. J Neurosci. 2015 Apr 1;35(13):5385-96. doi: 10.1523/JNEUROSCI.4880-14.2015.

The main Deliverables of this project, the review (manuscript above) and the human phase connectome publication(s) (partial data presented above) are *in preparation* at the time of the writing of this report.

Theoretical analysis project component investigating the principles underlying oscillatory network organization in the presence of time lags

Network couplings of oscillatory large-scale systems, such as the brain, have a space-time structure composed of connection strengths and signal transmission delays. As the spatial distribution of the connection strengths strongly influences the network dynamics, so does the spatial distribution of time delays. When the signal transmission delays are small with regard to the characteristic time scale of the system, then they can be mostly ignored. In the brain, the time delays of signal transmission (10 to 200 ms) are on the same scale as the signal operation (10 to 250 ms) (Buzsáki and Draguhn, 2004) and contribute critically to the system's spatiotemporal organization. Oscillatory network systems are particularly sensitive to changes in signal transmission delays, because shifts in phasings of the oscillators may easily change the nature of the mutual influences from excitatory to inhibitory and vice versa, affecting the overall synchronization behaviour of the network. The latter, is one of the hypothesized key mechanisms of brain function (Fries, 2005b; Varela et al., 2001) and cannot be ignored when studying large-scale networks. With the advance of non-invasive structural imaging techniques, large-scale network modeling approaches of the entire brain have now become feasible using biologically realistic connectivity, the so-called Connectome, and spatially distributed delays (Deco et al., 2009; Ghosh et al., 2008), Figure 32A.

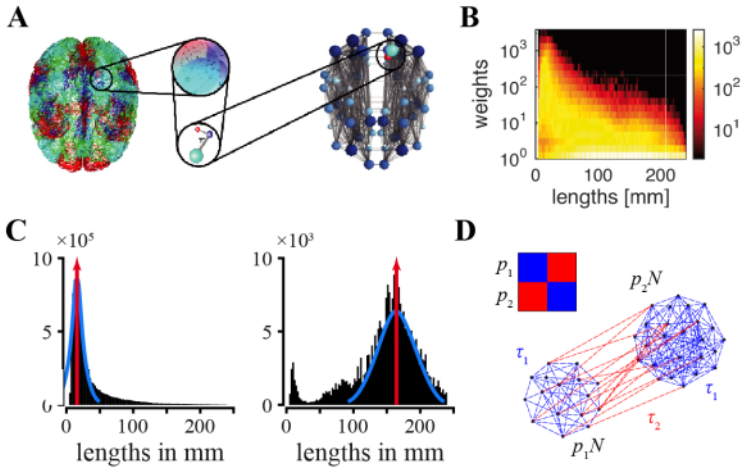


Figure 32

A) Two forms of connectivity are implemented in large-scale brain models: local connectivity (left) between near neighbors and global (right) representing the human Connectome. (B-C) Averaged tract lengths and weights from connectomes of 4 healthy subjects. Joint distribution B), and C) histogram of weighted lengths for intra and inter- hemisphere links. D) Sketch of the spatial delay-imposed structure of the brain used as a first approximation of its space-time organization.

Upon analysis of experimental connectivity data (Van Essen et al., 2013), the results imply that the lengths of connection routes between brain areas are multimodaly distributed (Figure 32B-C). This insight suggests that the complex space-time structure of the connectivity maybe approximated by a less complex mode decomposition, which will aid in the mathematical analysis of the large-scale brain dynamics.

We have provided a first theoretical framework, which allows treating the space-time structure of network couplings as a whole with regard to its effects upon network synchronization. By decomposing the spatial distribution of time delays into spatial patterns within the couplings' space-time structure, we could analytically compute the synchronization characteristics of the network. This analysis was performed on idealized

networks of phase oscillators with a connectivity structured by time-delays. We found that oscillators group in phase-shifted, anti-phase, steady, and non-stationary clusters as a function of the delay structure and computed their stability boundaries analytically for different frequencies. These synchronization patterns can be controlled by rearranging the time delays in the network. We thus clearly demonstrated that it is not just the connectivity that matters in oscillatory large-scale networks, as the brain, but time delays are of equal importance.

The principles of the modelling

We focus on the phase synchronization at different frequency ranges between distant regions of the brain. The aim of the model is to explain the mechanisms behind this phenomenon and to point the possible ranges and regimes of large-scale brain dynamic and consequently the phase differences between brain regions. Finally, the model should also encompass the empirically obtained results, based on the phase locking values (PLV) between electrophysiological measurements. PLV is a statistical measure for similarity between phases of two signals and although it can happen that the signals are coherent just by chance, in most of the situations it is assumed that a synchronization phenomenon is responsible for this coherence. Hence, statistical testing is necessary to distinguish these cases and statistically significant level of coherence to be obtained.

Since the experiments apply Hilbert or wavelet transform on band pass-filtered time-series for obtaining the phase at different frequency, the underlying dynamics of our model is assumed to consist of phase oscillators. Empirically only 1:1 synchronization is assumed, thus justifying our choice of modelling different frequency bands separately. Finally, in order to account for the non-deterministic nature of the brain dynamics, as observed in the experiment, white Gaussian noise was added. Its intensity was however kept much lower than the coupling strength, such that the main source of the heterogeneity in the model was the delayed interactions between oscillators. Still, the noise had significant influence to the dynamics at the microscopic level of each oscillator. By keeping the model simple, whilst accounting for the oscillatory nature of the brain, measured by PLV, we are able to directly apply the theoretical analysis that describes how the spatial distribution of time delays determines the synchronization of coupled oscillators.

It should be noted that there exists invariant scaling between the frequencies and the time-delays. For example increasing the transmission velocity over the fixed lengths as given by the connectome, has the same effect as decreasing the frequency. This however, only shifts the observed dynamical regimes at the frequency/time-delays parameters space. Hence by fixing the velocity at the realistic range of values, we can study the impact of the delays for different frequencies. We have performed simulations for velocities between 2 and 10 m/s, whilst in the results shown in the figures this is fixed at 5 m/s.

We used structural and diffusion MRI from randomly chosen healthy subjects data from the Human Connectome Project (Van Essen et al., 2013). The connectome was obtained for Desikan Killiany atlas (Desikan et al., 2006b) resulting in 68 cortical regions, where each link is described by a length and a weight that are averaged over all the tracts identified between those regions. Based on our analytical analysis this was then simplified, which proved to be sufficient to capture all the possible network dynamics of the realistic brain network. Namely, based on the tract analysis (Figure 32B-C), we decomposed the spatio-temporal connectome in two modes as a first approximation; (Figure 32D). The values of each of them are empirically obtained as the modes of the intra and interhemisphere distributions of the connectome links, Figure 32C. Nevertheless, this captured most of the synchronization profiles observed in the more realistic approximation and in the full network.

Modeling the experimental protocol

The global dynamics was described by the Kuramoto order parameter, which is divided in two spatial subnetworks following the space-time approximation. These modes, beside the incoherence, could be either in- or anti-phase, as predicted by the theoretical analysis. Hence for synchronization, most of the oscillators would have phase-lags around 0 or $\pm\pi$, depending on the frequency and the level of coupling/coherence. Besides the global (synchronization over nodes) we also analyzed the PLV (synchronization over links), providing direct link between the model and the experiment. Namely, time-windowed (TW) PLV was applied for each pair of nodes with length varying between 5 and 15 periods of the cycles (in the presented results it is set to 5, for consistency with the experimental protocol). However, for each of these cases a different level of significance was calculated based on the shuffled time-series and the uncoupled network with identical noise level. The former generally yielded lower values, meaning that the latter was adopted for most of the parameters range.

The simulations were run for at least 150 seconds after the transient period and identified the TW in which the PLV was significant, based on the both surrogate measures. From the complex PLV we obtained the phase difference in those TW. In addition, we also calculated the phase difference directly from the time-series of the phases at each oscillator. The circular statistics of both measurements was very similar, with the former representing more coarse-grained version of the latter. Similar holds for the levels of significance, the lower one was only making the peaks in the distribution of the phase differences to be flatter and less pronounced.

Results of the reduced model

The most important contribution of the model was the fact that it captures the statistics of the phase lags as observed in the experiment, whilst still accounting for the dynamical regimes predicted in the theoretical analysis.

In Figure 33, for fixed frequency the global dynamics changes from incoherence to anti- and in- phase coherence for increasing the coupling strength. However, even at global incoherent state, due to the heterogeneity of the connectome, local, link-wise coherence is observed. This is reflected at the matrix showing the circular average of the phase lags during the TW of significant PLV for each pair of regions. There are no empty entries in this matrix, meaning that even though the global level of coherence and the long-time averaged PLV are both very low, still there are short periods of time where the PLV is higher than the level of the surrogates. These are probably result of the stochastic nature of the dynamics at very low coupling, since there is no spatial structure, and the phase-lags are flatly distributed with relatively high variance.

By increasing the coupling, the coherence between links increases, and nodes organize in the space resulting in peaks of phase-lags at 0 and $\pm\pi$. The stronger connected nodes (with higher in-strength) are more strictly following this division: intrahemisphere links are in-phase (phase-lags ~ 0), and interhemisphere links are anti-phase. The links between the weaker nodes on contrary still have flatly distributed lags. This is seen from the matrices sorted by increasing in-strength of the nodes, i.e., the sum of all the weights for that node.

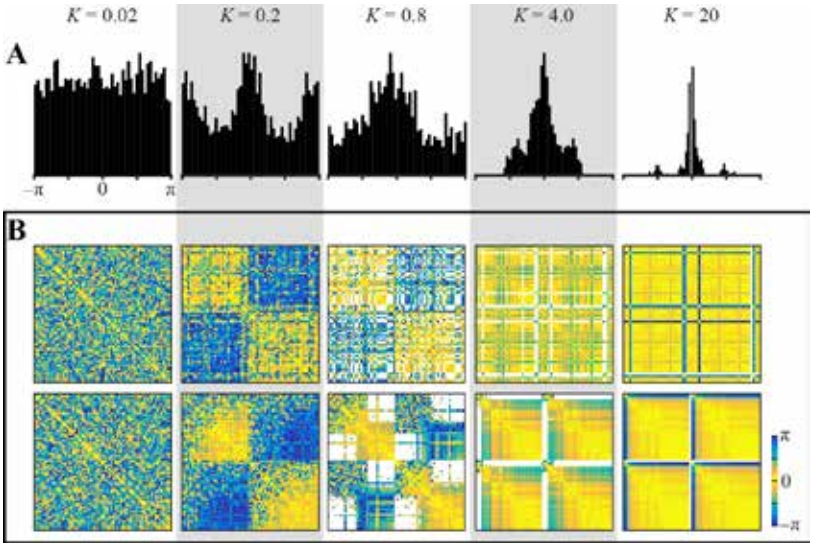


Figure 33: Regimes of the large-scale network dynamics, as reflected in the phase distribution between the nodes. Coupling strengths are increasing for fixed frequency (f = 20 Hz), resulting in switching from global incoherence to alternating in- and anti-phase coherence.

A) Histogram of the phase differences between significantly coherent links. B) Spatial distribution of the phase difference; white corresponds to the links where no significant PLV was detected at any TW. Nodes in the upper row are sorted according to the Desikian atlas, and in the lower row they are sorted by the in-strength of the nodes, within the hemispheres. The latter unveils a structural organization, which beside the spatial distribution takes into account the strength of the nodes.

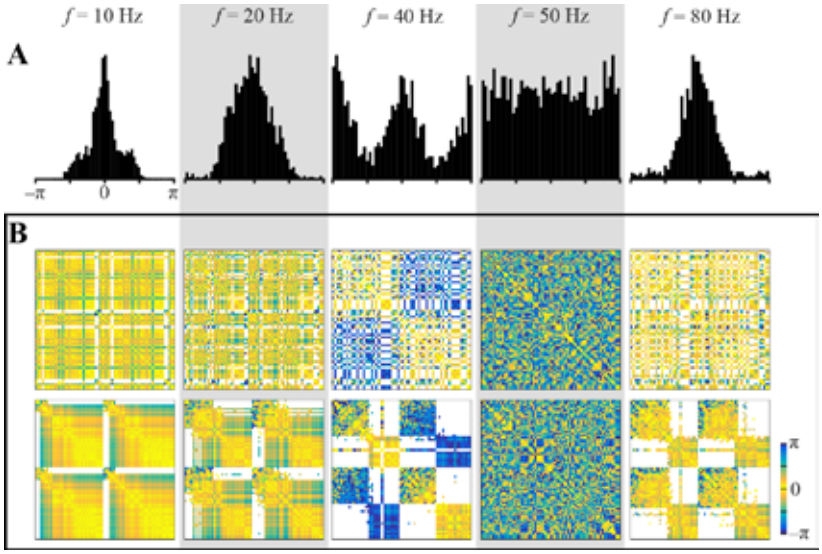


Figure 34: Large-scale network dynamics for different frequencies/delays. Natural frequencies are increasing for constant coupling (K=2) and noise strength, resulting in alternating switching from in- to anti-phase coherence, but also to incoherence (anti-resonance phenomenon of the time-delays).

(A) Histogram of the phase differences between significantly coherent links. (B) Spatial distribution of the phase difference; white corresponds to the links where no significant PLV was detected at any TW.

For higher coupling reorganization occurs. Most of the nodes become in-phase, but some of them are completely incoherent with others. These behave as detached, besides the global

coupling being much larger than the one needed to observe much higher static PLV and periods of significant PLV. These are in general links between very strong and very weak nodes, both intra and interhemisphere. So whilst the strong nodes are in- or anti- phase locked between each other depending in which hemisphere they are, and weak nodes are weakly non-stationary synchronized between each other at some short periods of time. Finally, for very large couplings, most of the nodes become strongly synchronized, leading to 0 peaked phase lags. However, even here the links between very weak nodes have phase-lags around ±π/2.

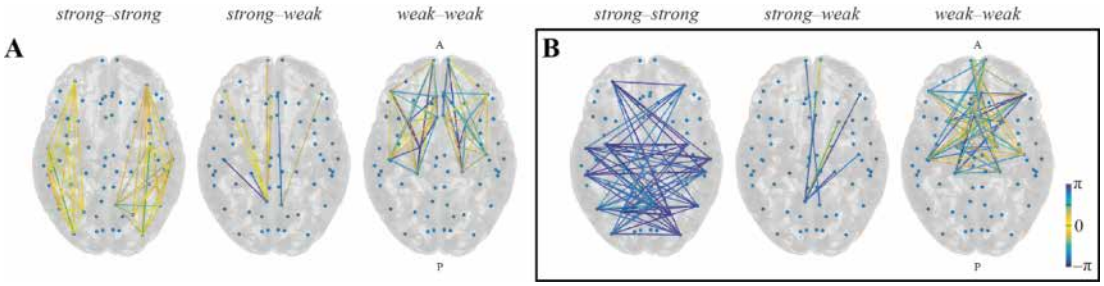


Figure 35: A) Intra- and B) inter- hemisphere subnetworks based on the strengths of the nodes, during the anti-phase synchronization for f = 40 Hz shown in Figure 34.

Each network consists of the eight strongest/weakest nodes at each hemisphere, corresponding to the spatial organization in the sorted matrix in Figure 34. The links between strong nodes within the same hemisphere have phase lags around 0, while between the hemispheres those are around ±π. The strong-link combinations have very few links with significant coherence, and weak-weak links have phase lags without a distinctive peak.

It should be noted that by increasing the significance level even higher (e.g. between 0.8-0.9), the smaller peaks, together with the flatly distributed delays at lower coupling strengths, become insignificant, yielding single or double peaked distributions around 0 and ±π. Thus, by simple readjusting the level of significant coherence, or the length of the time-windows, the model can account for different methodological experiments.

Due to the invariance to scaling of the time, the relative value of the time-delays compared to the period of the oscillations determines the synchronization. Thus, increasing the frequencies for constant transmission velocities impacts the coherence and the spatial organization of the oscillators. The anti-resonance phenomenon of the time-delays, which was theoretically predicted, also occurs in the connectome-based model. In Figure 34 the network dynamics alternates between in- and anti- phase, imposing spatial reorganization in the brain. Similarly, the periodicity of the delays/frequencies is also visible: after becoming bimodal at 40 Hz, the phase lags become flatly distributed due to the global incoherence at 50 Hz, before they return to the unimodal 0 peaked distribution, at 80 Hz similar as for low frequencies, <20 Hz.

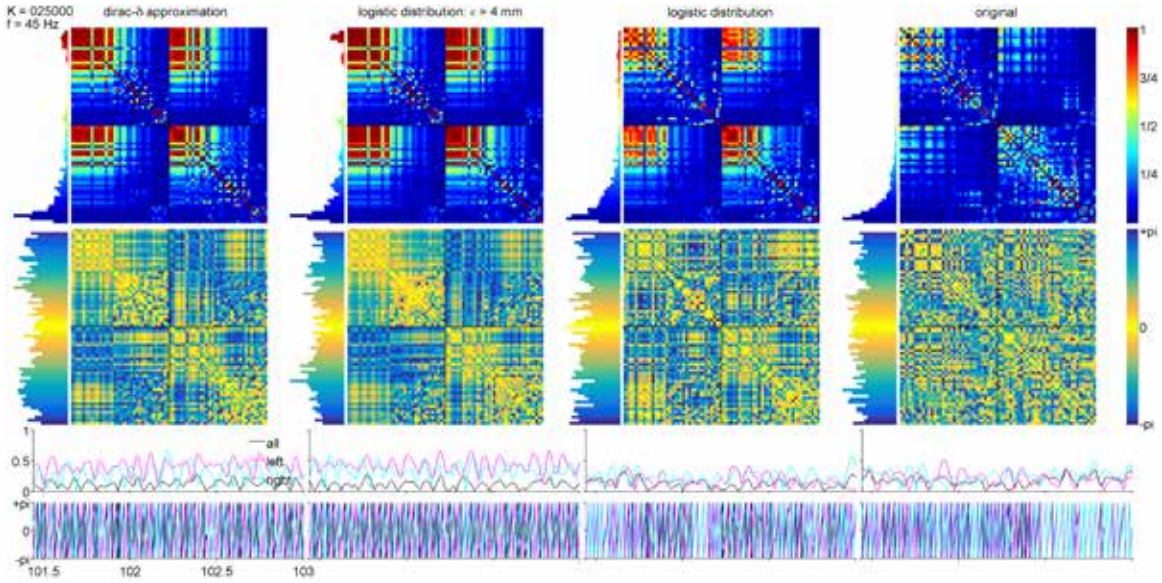


Figure 36: Comparison (from left to right) between the dynamical repertoires of the bimodal-delta reduction, logistic distribution with fixed width, best-fit logistic distribution and the full connectome.

Upper and the middle row: sorted by strength and matrix of the static phase lags, respectively, with the overall color-coded histogram. Bottom row: time series of the order parameter and its mean phase for the global network as well as the left and the right hemisphere.

The same organization based on the strengths of the nodes is also visible during coherence. The within hemisphere links between strong nodes have 0 phase lags and between hemispheres are either around 0 or around ±π (Figure 34). The links between weaker nodes on contrary are usually flatly distributed, while for the links between weak and strong nodes the significant PLV barely occurs for most of the frequencies. These distributions of the links between the anatomical regions are illustrated in Figure 35, for eight of the strongest and weakest nodes in each hemisphere (out of 34 in total). The in-strength based weak nodes are mostly in the frontal lobe and they form a subnetwork with links that have flatly distributed phase lags. The strong intrahemishere links mostly connect frontal and the occipital lobes and these are expected to have stronger coherence with phase lags around 0, while intrehemisphere links between those nodes with high in-strength of the weights have phase lags distributed around ±π.

Comparison between the reduced model and full connectome

Besides the extensive analysis of the reduced model for parameters (frequencies, coupling, transmission velocity) at realistic range, we also compared its dynamical repertoires with similar model but on full connectome, and two other more realistic approximations. First of these were bimodal distributions accounting for the same modes as the delta-peaks approximation, but with non-zero width. The second were more realistic fits of the actual distribution within and between the hemispheres (Figure 32D). The comparison was performed over different frequency bands and global coupling, and the results are shown in Figure 36. In order to focus on the importance of the different network topologies, these analyses were performed without adding noise to the dynamics and PLV and the phase lags are obtained only at the whole time-series.

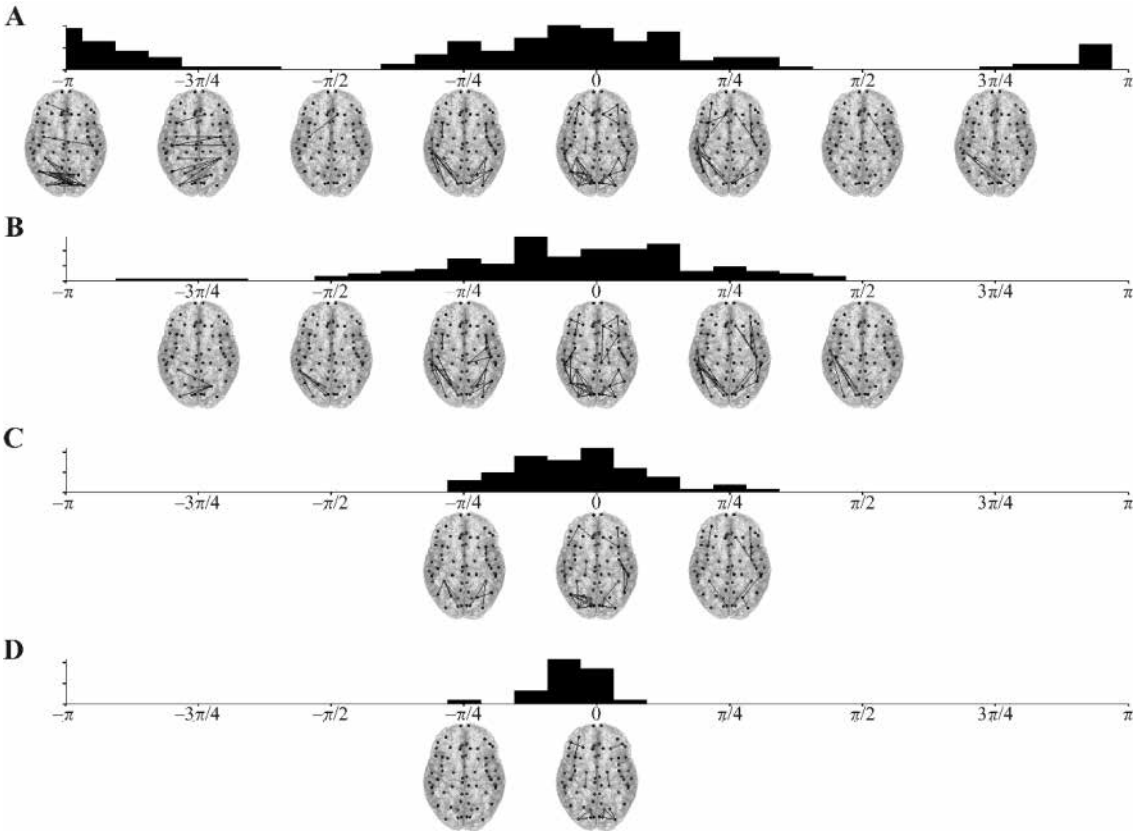


Figure 37: Phase lag spatial distribution over the brain links showing high phase synchrony

for A) the bimodal-delta reduction, B) the logistic distribution with fixed width, C) best-fit logistic distribution, and D) the full connectome. Spatial structure of the different phase lags amongst brain regions is shown on the large-scale brain network.

The results show that for higher frequencies and lower couplings, the heterogeneity of the full connectome prevents the spatial organization (intra-hemispheric synchronization in anti-phase, as observed in the time series of the order parameters, but also the static PLV value and phase lags). Thus, higher coupling is needed to compensate for the additional heterogeneity of the connectome, for achieving similar large-scale dynamics.

The spatial structure of the networks defined by the different phase-lag of their links is also analysed (Figure 37). The [-π, π] range is divided in 8 equal bins and networks are constructed based on the binning of the phase lags of the links. This is performed across different frequencies and couplings for full connectome and the three approximations. Results show that the zero phase lags always correspond to local (intrahemisphere) links and they appear for all the networks at all the frequencies and couplings. The differences between the networks appear in the organization of the intrahemisphere links, which are usually around ±π, and depending on the parameters space could be more prominent either in the full connectome or in the simplest delta approximation.

Discussion

We have here provided a first theoretical framework, which allows treating the space-time structure of network couplings as a whole with regard to its effects upon network synchronization. By decomposing the spatial distribution of time delays into spatial patterns within the couplings' space-time structure, we could analytically compute the synchronization characteristics of the network. This analysis was performed on idealized networks of phase oscillators with a connectivity structured by time-delays. We found that

oscillators group in phase-shifted, anti-phase, steady, and non-stationary clusters as a function of the delay structure and computed their stability boundaries analytically for different frequencies. These synchronization patterns can be controlled by rearranging the time delays in the network, while maintaining the connectivity identical. We thus clearly demonstrated that it is not just the connectivity that matters in oscillatory large-scale networks, as the brain, but time delays are of equal importance.

List of publications.

Petkoski S, Spiegler A, Prois T, Temprado JJ, Jirsa VK (2016) Delay-imposed structure on network dynamics of oscillators. Under review in Physical Review Letters

1.5 Visual Recognition

Task T3.1.1 - Rafael Malach (WIS), Stanislas Dehaene (CEA), Clément Moutard (CEA)

Introduction

Mental representations and mental models of visual objects are thought to play an important role in visual cognition. Starting from a retinal projection, the visual system is thought to construct an invariant representation of objects that also enables simulation, anticipation or problem solving. Even though these properties can be multi-modal, their visual implementation – and notably visual imagery – has been thought to be particularly central to human abilities, and has focused interest as soon as the end of the 19th century. For instance, the sensory physiologist Hermann von Helmholtz stated in 1894: "equipped with the awareness of the physical form of an object, we can clearly imagine all of the perspective images which we may expect upon viewing from this or that side, and we are immediately disturbed when such an image does not correspond to our expectations" (translated in (Warren and Warren, 1968) pp 252). Thus, visual recognition implies the existence of an internal model capable of assembling together, in a coherent manner, the collection of the different possible views of an object. Furthermore, this object model must be capable of being dynamically manipulated in a purely mental manner, in the absence of any physical stimulation, for instance through mental rotation (Shepard and Metzler, 1971). Finally, it can be spontaneously recalled or evoked during spontaneous thoughts.

While visual recognition is extensively investigated, the high-level representation of internal models of objects, detached from direct perception, is rarely studied. In the present task, therefore, we deliberately addressed this high-level of representation. First we reviewed what is currently known about the "ignition" of large-scale brain networks that supports conscious representation and mental manipulation during perception, recall, and spontaneous thought; and then collected new data sets on the links between perception and mental imagery (data set 1) and on the relations between brain activity during movie watching and spontaneous resting-state activity (data set 2).

Review of the cortical architectures underlying spontaneous fluctuations and non-linear ignitions

Clément Moutard, Stanislas Dehaene, Rafael Malach "Spontaneous Fluctuations and Non-linear Ignitions: Two Dynamic Faces of Cortical Recurrent Loops", Neuron, Volume 88, Issue 1, p194–206, 7 October 2015

Abstract

Recent human neurophysiological recordings have uncovered two fundamental modes of cerebral cortex activity with distinct dynamics: an active mode characterized by a rapid and sustained activity ("ignition"), and a spontaneous (resting state) mode, manifesting ultra-slow fluctuations of low amplitude. We propose that both dynamics reflect two faces of the same recurrent loop mechanism: an integration device that accumulates ongoing stochastic activity and, either spontaneously or in a task-driven manner, crosses a dynamic threshold and ignites, leading to content-specific awareness. The hypothesis can explain a rich set of behavioural and neuronal phenomena, such as the perceptual threshold, the high non-linearity of visual responses, the subliminal nature of spontaneous activity fluctuations, and the slow activity buildup anticipating spontaneous behaviour (e.g. readiness potential). Further elaborations of this unified scheme, such as a cascade of integrators with different ignition thresholds or multi-stable states, can account for additional complexities in the repertoire of human cortical dynamics.

Data set 1: MEEG recordings of the time course and manipulation of view-specific and view-independent models of objects

The present project aims at collecting detailed magneto-encephalograpy (MEG) data on predictive mental models and the brain dynamics associated to visual perception and visual imagery of spinning objects. Notably, the following key questions concerning mental models in the visual modality were addressed: can we identify two successive perceptual stages in the human brain, one view-specific and the other one view-independent? Can humans learn to attach a new view to an internal model of an object, and build a mental model binding these two levels? Can we follow the mental representation of an object through time, both during overt perception and during a purely mental manipulation exploiting visual imagery? And how are mental models used to predict the upcoming state of an object?

Methods

The experiment systematically recorded human MEG and EEG signals, respectively from 306 and 60 sensors, while 20 participants were watching stimuli on a screen, following a sequence of different paradigms as described below. To include the interplay of both levels of mental representations, the stimuli were designed to provided different views of two objects (3D cartoons of a head and of a coffee-maker) both having their own view-independent identity, and their respective view-specific poses corresponding to a rotation in depth (from 0° to 360°). Because faces (and to a lesser extent coffee machines) are common objects for humans, mental models for these two categories are probably already available, arising from previous learning of the statistical perceptual regularities of this category. In order to study the learning of a novel view of these objects, a non-natural texture was added on the back face of each object: a checkerboard behind the head, and a colorful wallpaper of flowers behind the coffee-maker (for half of the subjects; and the inverse texture-object association for the other half). Thus, subjects needed to discover and learn a complete model of each object: front and profile views conforming to everyday life, and a back view with an unusual pattern.

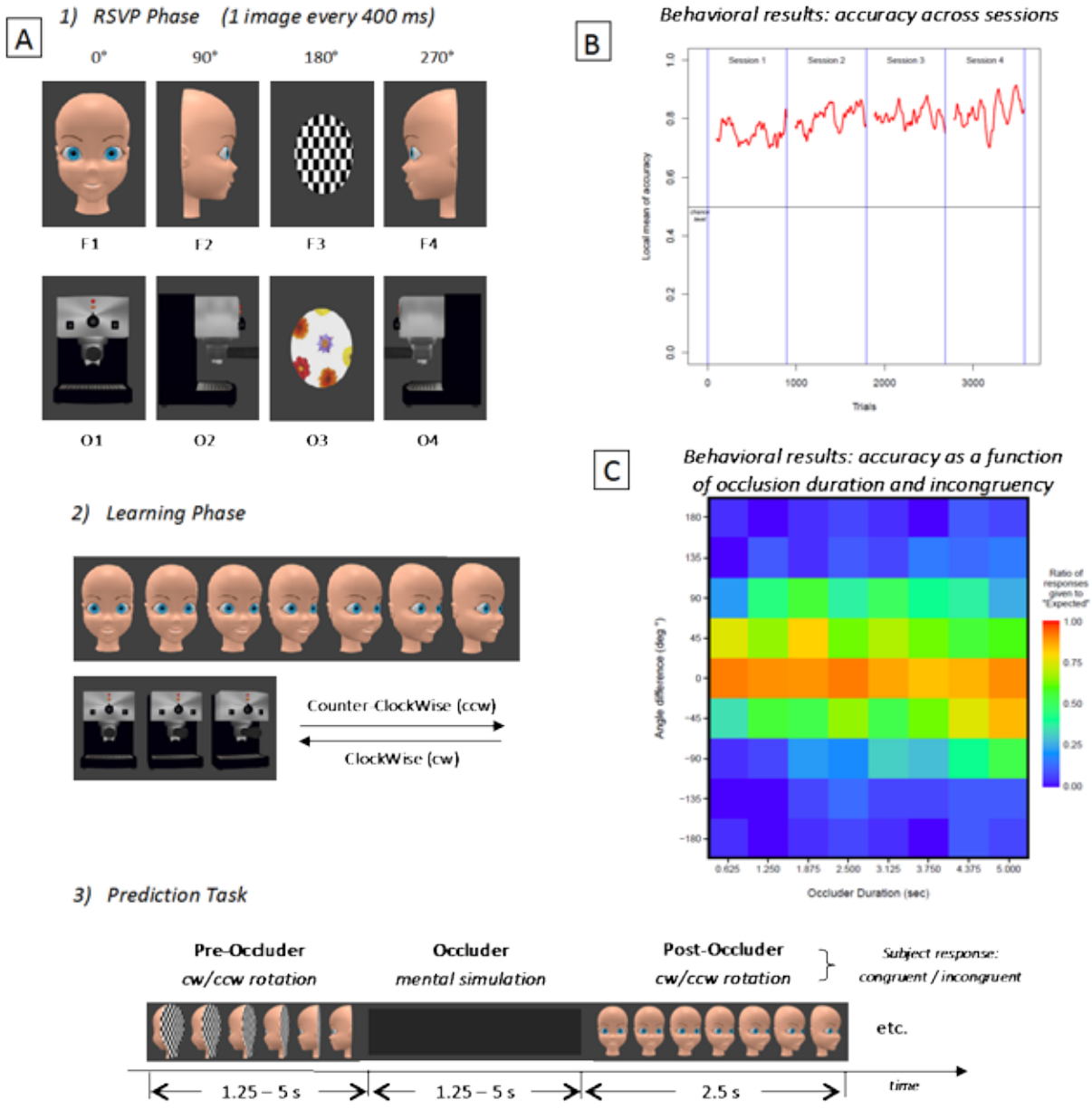


Figure 38: Experiment Design & Behavioural results

A) Experimental design comprising three different experimental paradigms: (1) rapid serial visual presentation (RSVP) of the 4 views of the 2 objects; (2) Learning by visual exposure to the rotating objects; (3) prediction task where subjects maintain a mental image of the rotating object throughout an occlusion period. B) Accuracy over a 100-trials sliding window of one subject in the behavioural version of the prediction task across the 4 sessions. The subject performed well above chance (accuracy>75%). C) Fraction of responses that the post-occlusion figure was congruent with the pre-occlusion figure, as a function of occlusion duration (x-axis) and degree of angular deviation (angular difference between the correct view and the one that was presented). The responses are "tuned" around the correct view, indicating that subjects were able to generate an accurate prediction of the outcome of the occlusion period. Angular precision decreases only slightly with occlusion duration.

The protocol was divided into 4 different phases. All the raw data (each phase, each subject) were acquired at Neurospin (CEA, Saclay – France) and can be found at http://sp3.s3.data.kit.edu/3_1_1/MEG_PredictiveVisualInternalModel/Raw/ accompanied by a text document named "readme.pdf" describing them. This data set is complete. The

first of the 4 phases, using a Rapid Serial Visual Presentation (RSVP) paradigm, aimed to provide detailed information about the perceptual stages of each view of each object. Eight images (the four cardinal views of both objects: front and back views, left and right profiles) were displayed for 100ms at a rate of 2.5Hz, each of them being repeated 60 times in a random order. The RSVP block was presented before and after extensive learning of the object model, with the goal to evaluate whether the novel view was eventually integrated into the internal object model. To eliminate shape cues, which may have allowed subjects to determine the assignment of views to objects even prior to learning, the back view textures were presented in a disc instead of with the shape of the object (Figure 38A).

The second phase consisted in four sequences of slow rotation in depth at 12 rpm: each of the two object spinned successively in both rotation directions (clockwise – cw, and counterclockwise – ccw) for 30 periods. This phase enabled the participant to become familiar with the stimuli, and notably to associate the unusual textured back views with the relevant objects and their other views. It also allowed us to collect data on the brain's responses to a physically rotating object, and to ask whether we could decode the object identity and orientation.

During the third phase, while the object was rotating, an occluder hid the object for a variable duration (1/4, 1/2, 3/4 or a full turn). The occlusion could only start and end on one of the 4 cardinal views of the objects (F1-F2-F3-F4-O1-…-O4). The participant was asked to pursue the rotation mentally with the same velocity. This phase therefore allowed us to probe the brain correlates of the mental model of a rotating object. Using multivariate decoders (L2-penalized logistic regression) trained for each object and also for each orientation, and applying generalization of decoding, we probe the presence of a shared internal model of objects during overt (physical) and covert (mental) rotation.

After a few seconds, the screen dropped, revealing either the appropriate object or the wrong object, in either the appropriate rotation or the counter-rotation direction, and with either the appropriate angle or a different angle. The subject's task was to press one of two buttons to distinguish congruent trials (in which the appropriate object appeared in the expected view and rotation direction) from incongruent ones (any other cases). Great care was taken to ensure that the numerous independent variables were balanced and crossed. Presenting objects with an incongruent view or identity provide a test of the predictive nature of the models on the two levels of representation (view-specific and view-independent). If the predictive coding hypothesis is correct, each of those mismatches should elicit error signals detectable via MEEG.

Finally, the fourth phase was fully identical to the first RSVP phase. The only difference relied on the learning experience of the participant. If the subjects acquired a model of the objects in the previous phases, including the unusual textures associated with each object, the presentation of the texture within a disc would elicit the activation of the view-invariant representation coding for the object identity. Thus, this design enabled us to test the integration of a novel view to an internal model of objects.

To obtain detailed psychophysical data, a behavioural version of the 3rd phase was also run outside of MEG. To improve the resolution of the psychometric measures, there were in this version 8 different occlusion durations (from 1/8 to a full turn) and the occluder could start and end on one of 8 different poses of the object (0°, 45°, 90°, 135°, …, 315°). Five subjects were tested during 4 sessions of 2.5 hours. Two subjects were rejected because they responded at chance level to the task of post-occlusion congruency determination. Figure 38B shows that subjects, when performing the task, were really able to mentally rotate the two objects. As depicted in Figure 38C, their accuracy was negatively correlated to the occlusion duration, and positively to the intensity of the incongruency (difference between the expected angle of view and the actual one). No

other factor was found.

Results and Discussion

The RSVP phases 1 (before learning) and 4 (after learning) with exactly the same stimuli enabled us to test if any association was made between textures and objects. Indeed, if it was the case, after learning the internal models, the textures should evoke the activation of the view-invariant information associated to an object (e.g. the checkerboard would elicit the face's identity representation). Thus, two decoders of the identity of the objects based on their natural views (F1/F2/F4 versus O1/O2/O4) were trained for each phase, and tested on their ability to generalize to the novel views.

Decoding of a static object: view-specific and view-invariant stages

Figure 39 demonstrates the high decodability of the identity of the object in both phases (first column). These decoders were then applied to the back view (F3 vs. O3): according to the hypothesis, one should expect no information concerning the object's identity before learning, but on the contrary its generalization from familiar to unusual views after learning. However, no such generalization was observed (Figure 39, second column).

One possibility is that no view-invariant information is processed by the brain. The next analysis was designed to test the latter possibility. To probe the existence of view-invariant content in the MEEG signal, four view-specific classifiers were built on the RSVP data to distinguish pair-wise poses of the two objects (e.g. F1 vs. O1). These pairs of view are correctly classified with a high accuracy as depicted in the matrices on the diagonal in Figure 40, and in the brown curves in the bottom row and in the right column. Each of these classifiers can also be applied to the other three pair-wise views to check the generalizability of the classification, and in which extent its nature is view-invariant. As shown in the matrices off the diagonal in Figure 40, the classifiers also generalized well to any other pairs of views, with the sole exception of the back view (in agreement with the above results, see above). While the peak of view-specific classification was reached around 85ms, the one associated to generalization – a.k.a. view-invariant – peaked around 140ms. Such delay rule out the possibility that this may be explained by low-level features alone. Thus, the evidence indicates the presence of two successive stages, view specific and view invariant, corroborating the hypothesis formulated in the introduction.

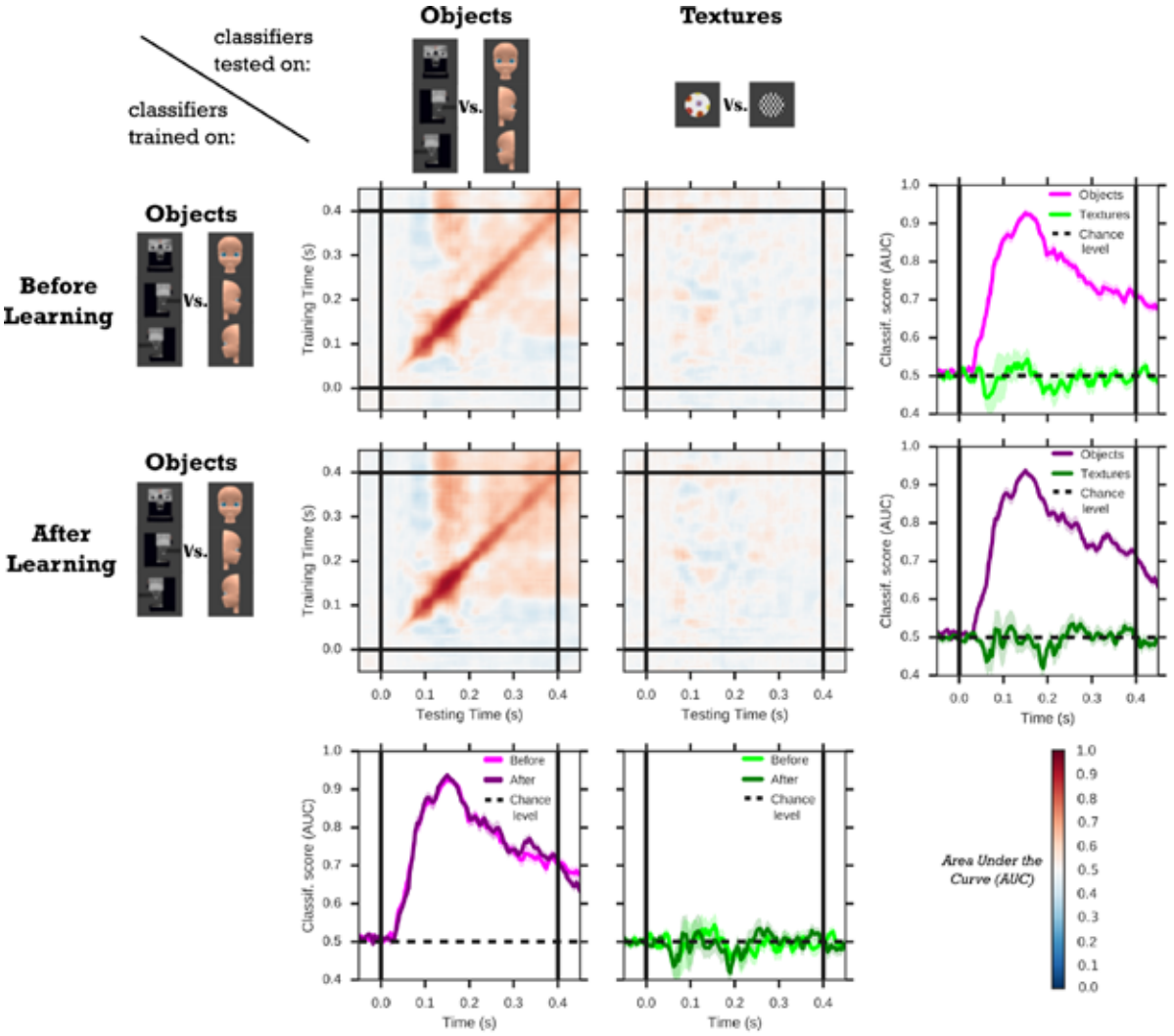


Figure 39: Decoding the identity of objects and generalizing to unusual views

Classification of the identity of the objects in the RSVP data set. Identity decoders were trained indistinctively on all objects' familiar views acquired before the learning phase (first row, i.e. upper two matrices) and after the learning phase (second row). For each time point from -50ms before to +450ms after picture onset, a distinct classifier was trained (training time = y-axis in the generalization matrices). It was then tested on each time point to evaluate generalization across time (testing time = x-axis in the generalization matrices). For each pair of training time/testing time, the performance of the associated classifier is calculated with the Area Under the Curve (AUC), and is plotted in the generalization matrices. The diagonal of each matrix is shown in the lower row and in the right column. While identity is easily decoded on familiar views, both before and after learning, there was no generalization to the unusual view in either case.

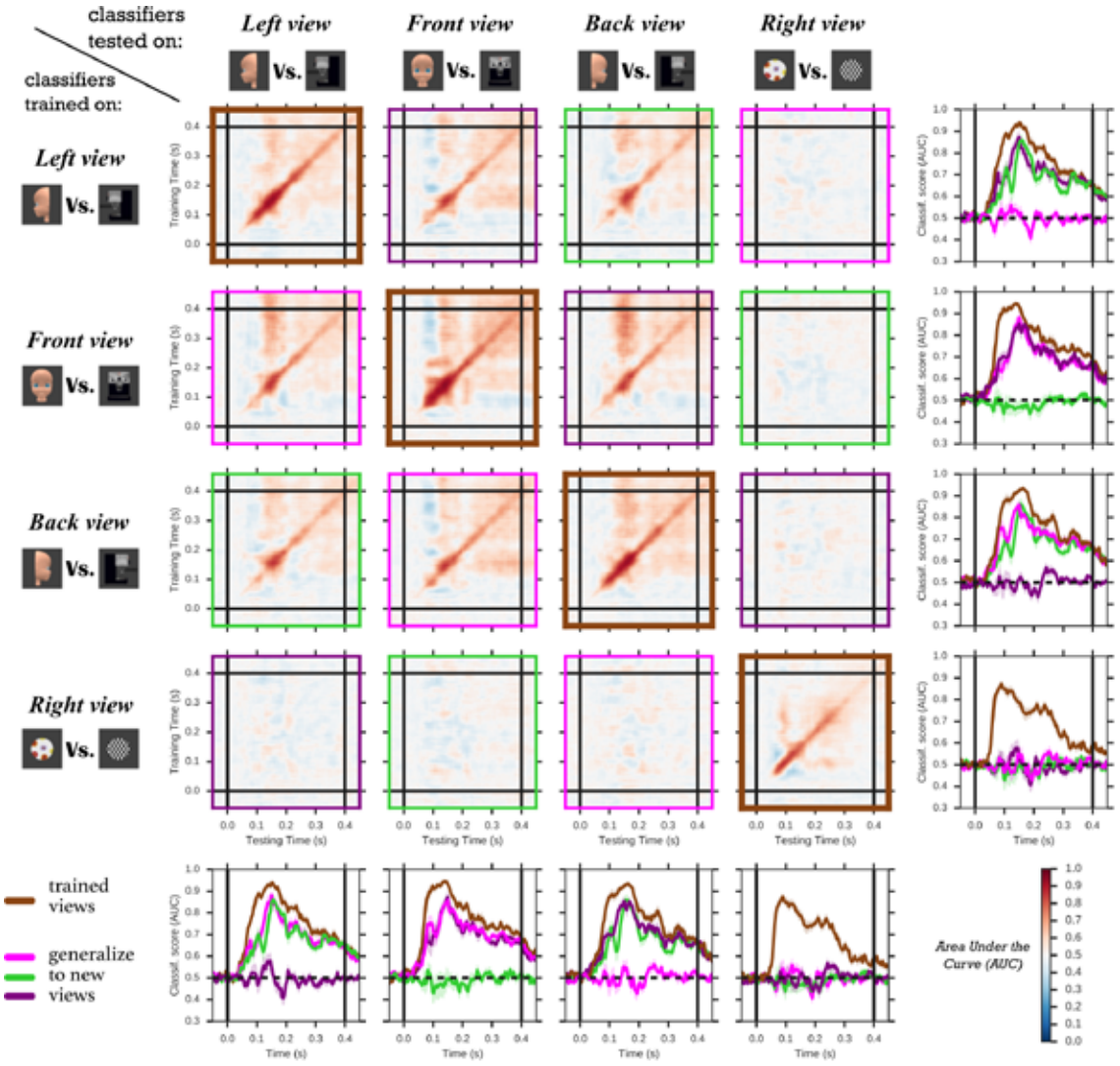


Figure 40: From view-specific to view-independent representations

Similarly to Figure 39, 16 generalization matrices are presented with their respective color-coded diagonals in the bottom row and right column. Each of the four rows corresponds to a pair-wise view-specific classifier (see pictures of the pairs of views), which is then applied to each of the same 4 pairs (four first columns). As shown by the matrices, classifiers do generalize from one familiar view to another familiar one, but there is no generalization to the unfamiliar one. The diagonal curves show a delay in generalization, with an earlier onset of the view-specific decoding (brown curve) compared to the view-invariant decoding (above chance non-brown curves).

Unfortunately, the data provide no evidence for learning of a novel view. Two possibilities may explain this negative result. First, the experiment in total lasted 2h30: maybe the neural encoding of new models need more time, and particularly to be consolidated during the sleep. However, other studies have shown that 1 hour is sufficient for a neuron to become sensitive to two different views and associate them (Li and DiCarlo, 2008). Another explanation is that the association is effective, but the related signal is too weak to be measured.

Decoding of a rotating object

The previous analyses were based on the static stimuli of the RSVP. The following ones will focus on the dynamic manipulation of the objects, during either overt or mental rotation. Some classifiers were trained to decode either the view-independent identity of the objects during overt rotation, or one of their view-specific pair-wise poses for all the angle of views (from 0° to 360°). This procedure was replicated during the mental rotation. Despite a successful decoding of the view-specific information in the physical rotation, no generalization across views was observed. Coherently, the decoder of the view-invariant identity of the objects was very low. Furthermore, no decoders were able to decode neither the identity nor the specific views of the objects that were mentally rotated.

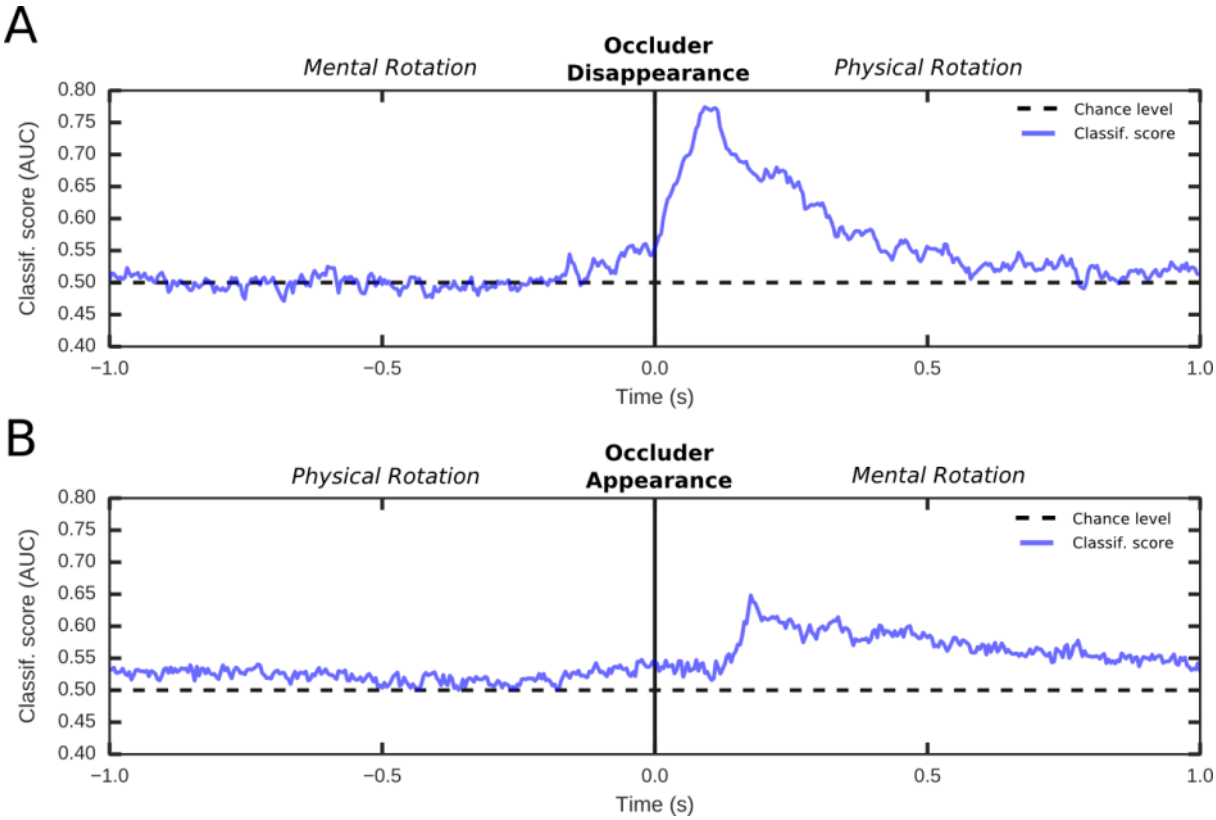


Figure 41: Transient decoding of object identity just after the beginning and the end of the occlusion period

A) Decoding object identity around the end of the occlusion period. Before the occluder disappears (left part of the curve), decoding remains at chance level, but the performance of the decoder rises suddenly and transiently upon reappearance of the object (end of the occlusion period). B) Decoding of object identity around the beginning of the occlusion period: again, decoding performance increases sharply with the transient, although in this case the object is not even present on screen.

Remarkably, however, further analyses found that it was locally possible to decode the identity of the object at two special times: just after the occluder appearance (Figure 41A) and just after its disappearance (Figure 41B). Furthermore, during the latter stage, we could also decode the congruity or incongruity of the view that appeared at the end of the occlusion period. This finding provides direct evidence that an internal model of the object was being sustained mentally throughout the occlusion period (in agreement with subjects' reports) and that its occasional mismatch with the view that eventually appeared led to a measurable error signal

Conclusion

This experiment led to mixed outcomes. On the negative side, in spite of major efforts, we could not detect any evidence of the acquisition of a novel view of an object, nor decode the mental model of a continuously rotating object, even when it was physically present on screen. However, we discovered that *transient changes* in this object could be decoded. Indeed, the entire data set is fully compatible with the idea that *only transients (i.e. error or update signals)* could be identified in MEG signals. It is particularly striking that view-independent information about object identity could not be identified in brain responses when the object was physically seen, but could be retrieved just after its occlusion, while the object was no longer present on the screen.

Altogether, these results can be interpreted as follow: any unpredictable occurrence (such as a rapid flash of an image in the RSVP, or the appearance and disappearance of the occluder in the 3rd phase) evokes a highly decodable signal, which fits with a prediction error consistently with the frame of predictive-coding theory. However, any predictable view yields very little or no detectable brain activity, as the incoming signal is being entirely cancelled out by the prediction arising from the internal model.

While the presence of error signals gave clear evidence of an internal model, we could not yet identify how this mental model was encoded in brain signals. Nevertheless, a cue is provided by a similar result in the literature: at the beginning of a sound, a transient neuronal activity is seen, but if the sound remains stationary, the neural activity returns back to baseline (deCharms and Merzenich, 1996). During the subsequent stationary period, only a change in the coherence of brain signals correlates with the perception of this continuing sound. By analogy, we speculate that oscillatory or synchrony mechanisms provide a potential candidate for the brain's continuous and sustained representation of an internal model of objects. In the future, time-frequency analyses will be conducted to evaluate this intuition. Meanwhile, the present work (which will be submitted for publication in the next few months) provides a high-quality data set against which to test new hypotheses concerning the internal representation of visible and hidden objects.

Data set 2: The patterns of co-activation during natural sensory processing uncovered through resting state and naturalistic stimulation paradigms

The work of the Malach group has focused on uncovering the fine details of visual cognitive architecture of the human brain using fMRI. The approach was based on the discovery that in the absence of stimulus or task, cortical networks spontaneously generate rich and consistent coherent patterns (also termed "functional connectivity"). While originally assumed to reflect large scale anatomical pathways, more recent studies revealed that these spontaneous (resting state) patterns show a highly detailed and intricate patterning also at the level of individual cortical areas and systems. In particular, previous research has demonstrated that in human retinotopic visual cortex, these spontaneous patterns are organized according to eccentricity lines rather than merely linking neighboring cortical sites. However, the underlying source of these intricate organizations has remained elusive. We and others have previously hypothesized that the fundamental process that shapes the organization of the spontaneous patterns is their habitual activation during the daily life of the individual. One prediction of this hypothesis is that the spontaneously emerging patterns should be more similar to patterns of co-activations produced by naturalistic stimuli that may better capture the natural statistic of such daily activations compared to more conventional laboratory-controlled stimuli. Here we tested this hypothesis by comparing resting state connectivity patterns to those induced by free viewing of naturalistic stimuli (a repeated movie segment), as well as patterns produced by more standard retinotopic stimuli- such as expanding and contracting eccentricity rings, and rotating wedges , as well as patterns predicted by a simple anatomical distance measure. Our results show that the movie driven correlations showed a significantly higher similarity to the resting state patterns compared to either eccentricity or polar mapping stimuli. These results could not be accounted for by higher reliability of the movie driven responses. These results were duplicated when subjects engaged in an auditory beep detection task- arguing against intentional visual imagery of naturalistic stimuli as the underlying source of the spontaneous patterns. Our results demonstrate the power of understanding cognitive architectures through resting state connectivity – and extend this discovery to the domain of naturalistic stimuli- and hence illustrate that such cognitive architectures reflect ecological cognition. The results thus successfully fulfill the central aims of the proposed HBP project. The entire work has been published recently (Wilf et al., 2015).So below I presented a shorter summary, while full details can be obtained from the published manuscript.

Methods

The main experimental procedures employed are depicted in Figure 42.

Our basic paradigm consisted of two resting state periods with eyes closed, and two repetitions of a naturalistic movie segment. In addition participants conducted an auditory beep detection task (Figure 42A). All participants also took part in an additional session for standard retinotopic mapping using two presentations of rotating wedge and expanding ring stimuli (Figure 42A).

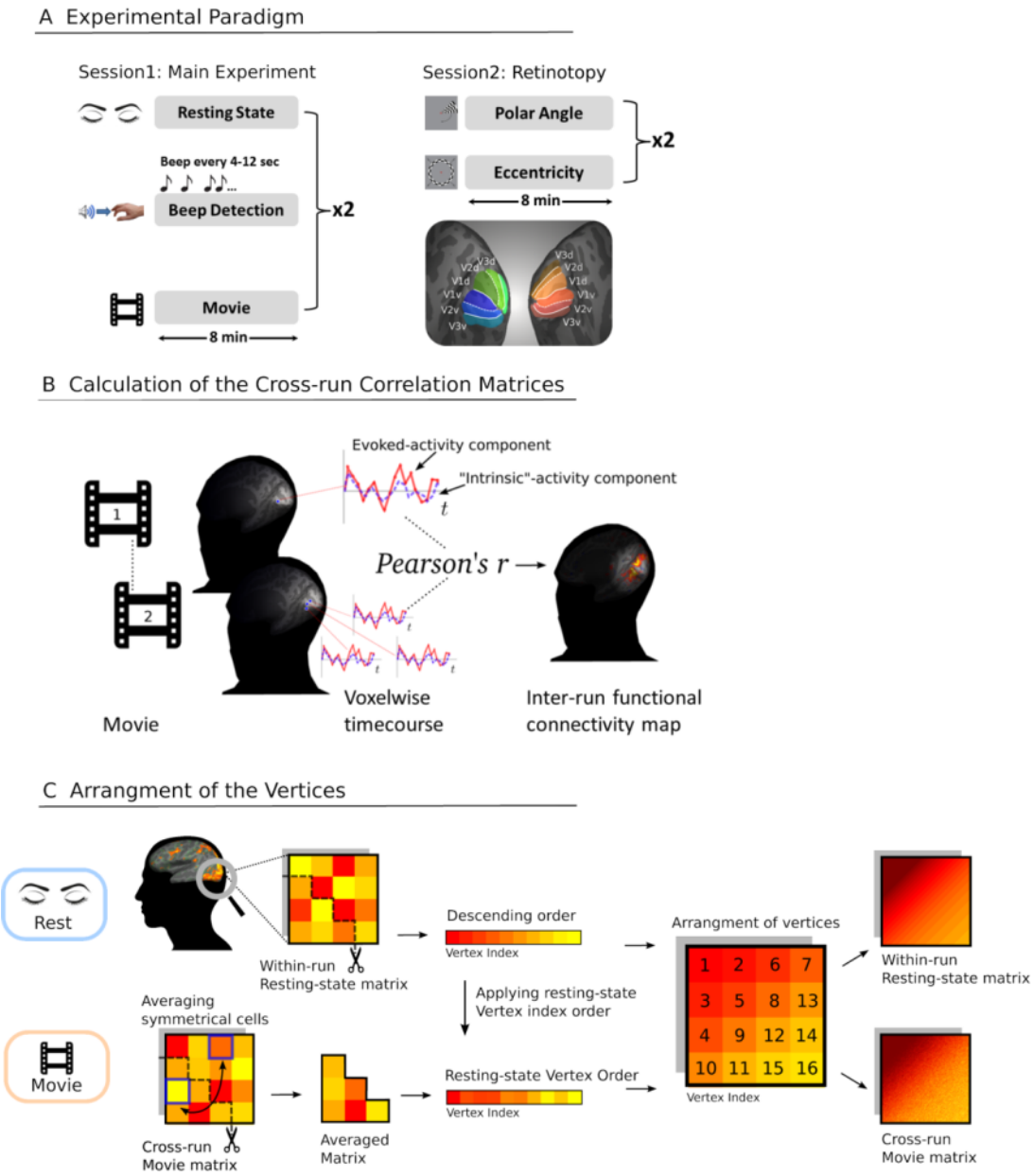


Figure 42: Experimental design and methodological approach

(A, left) The experiments in the main experimental session included resting state, auditory task and free viewing of a movie segment. Note that each condition had two repetitions, and the session always began with resting state. (A, right) In the retinotopic mapping session each visual condition was presented twice. These stimuli were used to define retinotopic visual areas for each participant (see methods). The order of the two sessions was counterbalanced between participants. (B) In order to avoid non stimulus-driven effects in analyzing visually-induced conditions, an inter-run approach was used. Stimulus driven functional connectivity was calculated by taking the seed voxel from one movie repetition and correlating it with all the other voxels in the other movie repetition (see methods). (C) Schematic illustration of the sorting procedure of the correlation matrices according to the resting state correlations (see methods).

Our results centered on quantifying the similarity between resting state patterns and the stimulus-driven patterns under all conditions (Figure 42). Figure 43 depicts a summary histogram of the similarity between the connectivity patterns under the different conditions. The Y axis depicts how similar the stimulus-driven pair-wise correlation pattern in each condition was to the pair-wise correlation pattern during rest, separately for areas

V1 to V3. Two aspects can be discerned: First, all stimulus-driven visual correlations showed a significant similarity to the resting state pattern (smallest $t(13) = 8.9$; $p < 0.001$ for all conditions ; one-sample t-test, Bonferroni corrected). Second, the movie-driven correlation pattern showed a significantly higher similarity to the resting state pattern compared to all other visual conditions. Overall, this similarity was significantly higher in V1 compared to the other regions (repeated measures ANOVA showed $F(2,26) = 5.04$, $p = 0.02$ for visual area; $F(3,39) = 38$, $p = 0.000005$ for condition; and no interaction; post-hoc tests show that the movie driven correlations were more similar to rest compared to all other conditions $p < 0.0001$). As a control, we computed a correlation matrix between resting state periods, assuming no consistent pattern should emerge. Indeed, this cross-rest condition failed to show a significant similarity to the within-rest condition, in accordance with the spontaneous nature of the fluctuations in each of the resting state runs. These effects reveal the striking correlation between the connectivity architecture during naturalistic conditions and resting state.

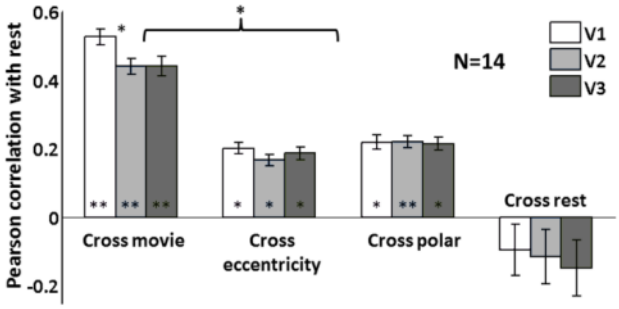


Figure 43: Naturalistic movie pattern shows significantly higher correlation to resting state pattern than other visual conditions.

All visual conditions showed significant correlation to resting state, but the movie produced significantly higher correlations than all other visual conditions. Cross-resting state matrix failed to show a significant correlation to the within run resting state patterns. Error bars denote standard error of the mean (±SEM); *$p < 0.05$ **$p < 0.000005$.

Discussion

Our results constitute a successful fulfilling of the aims as outlined in the HBP project for our group- namely to uncover the cognitive architecture of visual representations under conscious perceptual states. We have extended beyond these aims by demonstrating the tight link between such cognitive architecture and naturalistic, ecological vision. This work has now been published (Wilf et al., 2015).

More specifically, our findings show that the correlation patterns that emerge spontaneously in retinotopic visual cortex during resting state resemble those that are generated by naturalistic visual stimuli. The movie-driven patterns exhibited a significantly higher similarity to the spontaneous connectivity patterns compared to the patterns evoked by more conventional retinotopic mapping stimuli. Our study confirms and extends previous findings that support the notion that the resting state correlation patterns are not merely a reflection of large scale anatomical networks, but show additional intricate patterning at finer detail - i.e. within sensory systems and within individual visual areas as well.

It is important to emphasize that the main focus of the present study was on the type of visual stimuli that may induce activations that best mimic the spontaneously emerging

patterns. The space of such possible stimuli is obviously enormous and impossible to exhaust. We therefore opted to compare specifically three types of stimuli - two types that are commonly used in retinotopic mapping experiments (eccentricity rings and polar wedges) and one that was aimed to simulate naturalistic vision. We also compared the movie driven patterns to four putative organizational principles, in which the patterns followed eccentricity, polar, retinotopic location, surface distance and volume distance. In all these comparisons, the cross-movie correlations proved to be significantly more correlated to the resting state patterns.

However, we acknowledge that we cannot rule out the possibility that additional stimuli of different kinds may produce patterns of activations that will show yet higher similarity to the spontaneous connectivity patterns. Nevertheless, even within this rather limited range of principles, our results are robust and informative as we discuss below.

This result is significant in offering a potential solution to the more general problem of using the uncontrolled resting state paradigm. Thus, our results demonstrate that it is possible to employ controlled tasks even when examining spontaneous "resting state" connectivity patterns, as long as the networks under study are not driven by the task.

Thus, our results show that the naturalistic, free viewing conditions indeed induced visual activation patterns that were closely reflected in the patterns that emerged spontaneously in retinotopic visual areas, both during rest and during the auditory task. This is particularly remarkable because, the naturalistic paradigm, lacking eye movement controls and specific tasks, substantially degraded the reliability of the movie-induced patterns. It is likely that under more reproducible conditions, such as repeating the movie several times (at long intervals), the correlations between the spontaneous and movie-driven patterns could become even higher. Thus, our results clearly demonstrate that the spontaneous, stimulus free patterns, while showing significant similarity to co-activations driven by conventional visual stimuli such as eccentricity and polar mapping, nevertheless contained

The observed reflection of naturalistic movie patterns in the spontaneous connectivity patterns is nicely compatible with our proposed hypothesis (Harmelech and Malach 2013), that an important factor that shapes the resting state patterns is a Hebbian-like strengthening of functional connectivity, induced by habitual co-activations of cortical networks during daily life. Such co-activations presumably lead to synaptic changes that later modulate the correlation patterns appearing spontaneously during rest. Compatible with this suggestion, animal studies have shown that the similarity between naturalistic stimulation and spontaneous activity increases with development, while the similarity between non-naturalistic (moving grating) visual stimuli and spontaneous activity does not (Berkes et al. 2011). Here we propose that habitual visual experience plays a role in shaping cortical connectivity during adult life as well.

While our results are compatible with the notion that the spontaneous connectivity patterns capture aspects of naturalistic vision, this does not rule out the possibility that important aspects of the spontaneous connectivity patterns are guided by intrinsic factors that are not necessarily experience dependent.

Because of the rich, multi-scale and diverse nature of naturalistic stimuli, it is extremely difficult to uncover the full set of organization principles and statistical tendencies that characterize them. Hence it is difficult to point out what aspect of the activations endowed the movie-driven patterns with their higher similarity to the spontaneous connectivity patterns compared, to, for example, the eccentricity-driven correlations. Future studies, moving gradually from highly controlled, schematic stimuli such as the eccentricity rings to more naturalistic stimuli could help in answering this question.

Finally, it is important to caution, that although our study shows a highly significant correlation between movie-driven patterns and the spontaneously emerging ones, this correlation does not prove causation. At this point, we cannot rule out the possibility that the movie stimuli simply activated a complex of intrinsically connected structures that also appear in the spontaneous fluctuations. Examining individuals that engage in drastically different visual environments in their daily life may provide important clues to this question. Thus, studying individuals with retinal abnormalities, those that work in artificial lighting conditions or are exposed to long periods of outdoor, peripheral visual stimuli, could further test the notion that the environment shapes the patterns of correlations that emerge spontaneously in the absence of visual stimulation. To conclude, our results show a new complete study demonstrating a new link between resting state activity and naturalistic cognitive architecture. These results accomplish the set goal of the part of our research supported by HBP. The data has so far not been used by others.

A **Dataset Card Information** has been completed (see DIC Task T3.1.1 "Architecture of functional visual cognitive networks of the human brain").

Data Provenance and Location

All data were acquired by Meytal Wilf and Francesca Strappini at the Weizmann Institute Human brain Imaging center. Stimuli and Data are available at the Weizmann Institute of Science, Department of Neurobiology, Rehovot, Israel.

Publication

The data has been published in Cerebral Cortex:
Wilf M, Strappini F, Golan T, Hahamy A, Harel M, Malach R. "Spontaneously Emerging Patterns in Human Visual Cortex Reflect Responses to Naturalistic Sensory Stimuli." Cereb Cortex. 2015 Nov 15. pii: bhv275)

1.6 Circuits linking perceptions to actions

Task T3.1.2 - Martin Giese (EKUT)

Review of the cognitive architecture for action processing

Martin A. Giese, Giacomo Rizzolatti "Neural and Computational Mechanisms of Action Processing: Interaction between Visual and Motor Representations", Neuron, Volume 88, Issue 1, p167–180, 7 October 2015

Abstract

Action recognition has received an enormous interest in the field of neuroscience over the last two decades, with a strong impact also in many other disciplines such as philosophy and robotics. In spite of this interest and impressive numbers of publications on this topic, the knowledge in terms of fundamental neural mechanisms that provide constraints for underlying computations remains rather limited. This fact stands in contrast with a wide variety of speculative theories about how action recognition might work, and how it might interact with other cognitive brain functions. This review focuses on new fundamental electrophysiological results in monkeys, which provide constraints for the detailed underlying computations, where we focus particularly on mirror mechanisms and interactions between visual and motor processing. In addition, we review models for action recognition with concrete mathematical implementations, as opposed to purely conceptual models. We think that only such implemented models can be meaningfully linked quantitatively to physiological data and have a potential to narrow down the many possible computational explanations for action recognition. In addition, only concrete implementations allow to judge whether postulated computational solutions are feasible and can be implemented with real cortical neurons.

Data set: Benchmark data set constraining computational mechanisms for the recognition of goal-directed hand actions

We provide two data sets that have been essential for the development of a quantitative neural model for the recognition of goal-directed actions (Fleischer et al., 2012, 2013), and for our newest modelling work that links perceptual and motor representations of actions (Christensen et al., 2011). Data from monkeys will not be provided since the HBP decided not to support research on monkeys in the Ramp-Up Phase. As consequence, primate researchers in Tübingen decided not to provide any data. Data on imaging and MEG were never promised by the task leader (M. Giese). We deny any responsibility for these claims in the work program, nor do we have access to such data sets.

The provided data contains the following parts: a) psychophysical data from a study by (Christensen et al., 2011), testing the influence of time delays and spatial transformations between concurrent execution and visual feedback for actions on the visual detection performance for action stimuli. The data set contains data from three experiments, varying the delay and the spatial congruence of the visual feedback, and one control experiment where the correspondence between the observed and the executing arm in terms of the body side (right or left) was varied. Data are provided as Excel spreadsheet. b) Psychophysical data from an experiment (Fleischer et al., 2012) that compared naturalistic hand actions and causality stimuli consisting of moving discs. Both stimulus classes were parametrically modified in ways that reduce the perceived impression auf causality (adding shifts, rotations, delays, and pause intervals to the motion of the hand). Subjects reported the perceived causality impression and the naturalness of the stimuli on Likert scales. Data are provided as Excel spreadsheet. c) For the testing of computational vision and neural models we also provide a stimulus set form this experiment as videos. Shown actions are pushing and grasping, with different amounts of the manipulations for destroying causality that were described above.

Neural models for the recognition of goal-directed actions and its interactions with motor representations

The available funding in the HBP (equivalent to 1/2 PhD student) did not allow us to embed or adapt large-scale models for HBP simulation frameworks. Instead, we decided to focus on a small number of innovative theoretical problems that could be treated with the small available manpower. In addition, we focused on linking a simplified model for the coupling between perceptual and motor representations of actions with HBP tools for the simulation of large-scale spiking networks. This simplified model is made available as part of the HBP data sharing initiative. Until the end of the Ramp-Up Phase, we plan to extend this core model by a visual pathway that works on real images and an output stage that simulates full-body motor output. These additional stages will not be integrated in the HBP framework since the available manpower is not sufficient for the realization of the relevant interfacing work, and because EKUT is going to leave the HBP after the Ramp-Up Phase.

The work described in this Deliverable consists of three parts: a) extension of an existing model for action recognition for the treatment of multi-stability and the perception of different stimulus views, and development of new theoretical methods for its analysis; b) extension of existing model by a new pathway for the processing of depth cues derived from shading, and related psychophysical experiments; c) development of a new neurodynamical model with spiking neurons that captures interactions between visual and motor representations of body movements, and implementation in NEST simulator framework. The following gives a short description of these three parts.

Neurodynamical model for multistability and the perceptual organization effects including multiple views in action recognition

This work focused on the neural dynamics that underlies the representations of multiple views of actions and the temporal integration of visual information of action stimuli. Extending previous models (Fleischer et al., 2013; Giese and Poggio, 2003), we extended the core circuits, which represents temporal sequences of perceived body postures as travelling activity peak in a one-dimensional neural field, consisting of shape-selective snapshot neurons by extending it to a two-dimensional neural field whose dimensions encode the stimulus frame within the movie and the stimulus view. By devising an appropriate lateral interaction kernel we were able with this model the psychophysically observed multi-stability of action stimuli without depth cues (Vanrie et al., 2004), where perceptual switching occurs between different perceived stimulus views. In addition, the model contains adaptation processes whose parameters were fitted in detail to adaptation experiments in area IT since corresponding data from action selective neurons is largely absent. A simple version of this model was presented in (Giese, 2014). The actual version that reproduces multi-stability and adaptation effects of action perception, exploiting mechanisms that are also reproducing details of the much better studied adaptation effects in area IT, is defined by the following equation system:

$$
\begin{aligned}
\tau_u \dot{u}(\phi,\theta,t) &= -u(\phi,\theta,t) + w(\phi,\theta) * F(u(\phi,\theta,t)) + c_v v(\phi,\theta,t) \\
&\quad + s(\phi,\theta,t) + \xi(\phi,\theta,t) - \alpha a(\phi,\theta,t) \\
\tau_a \dot{a}(\phi,\theta,t) &= -a(\phi,\theta,t) + [u(\phi,\theta,t)]_+ \\
\tau_b \dot{b}(\phi,\theta,t) &= -b(\phi,\theta,t) + [r(\phi,\theta,t)]_+ \\
s(\phi,\theta,t) &= k(\phi,\theta) * (g(b(\phi,\theta,t)) r(\phi,\theta,t)) \\
\tau_s \dot{v}(\phi,\theta,t) &= -v(\phi,\theta,t) + [\dot{s}(\phi,\theta,t)]_+
\end{aligned}
$$

This model for the mean-field dynamics of body-shape selective neurons in action-selective cortical areas (e.g., the STS or area F5 in monkeys) contains stochastic fluctuations, two different adaptation processes (firing rate fatigue (FF) and input fatigue (IF)) in order to account for details of the relationship between adaptation and selectivity of neurons derived from data in area IT (De Baene and Vogels, 2010), and spike rate (SR) adaptation, accounting for he observed signal shape of real cortical neurons. The parameters of this model were carefully fitted to data sets from area IT (assuming that the underlying adaptation processes are similar the ones in action-selective areas) and to data from area F5 in premotor cortex. The model accounts with one parameter set for a variety of phenomena observed in electrophysiological experiments: (i) perceptual switching between different stimulus views for action stimuli without depth cues; (ii) size and time course of high-level adaptation effects in shape recognition, including subtle dependencies between stimulus selectivity and adaptation size which necessitate the assumption of a contribution of input fatigue (De Baene and Vogels, 2010); (iii) the result that action-selective neurons in area F5 show only very weak or even no adaptation effects for single stimulus repetitions (Caggiano et al., 2013; Kilner and Lemon, 2013).(iv)The model predicts a new stimulus type that should result in strongly increased adaptation effects in such action representations. (See (Giese, 2014) for further details.)

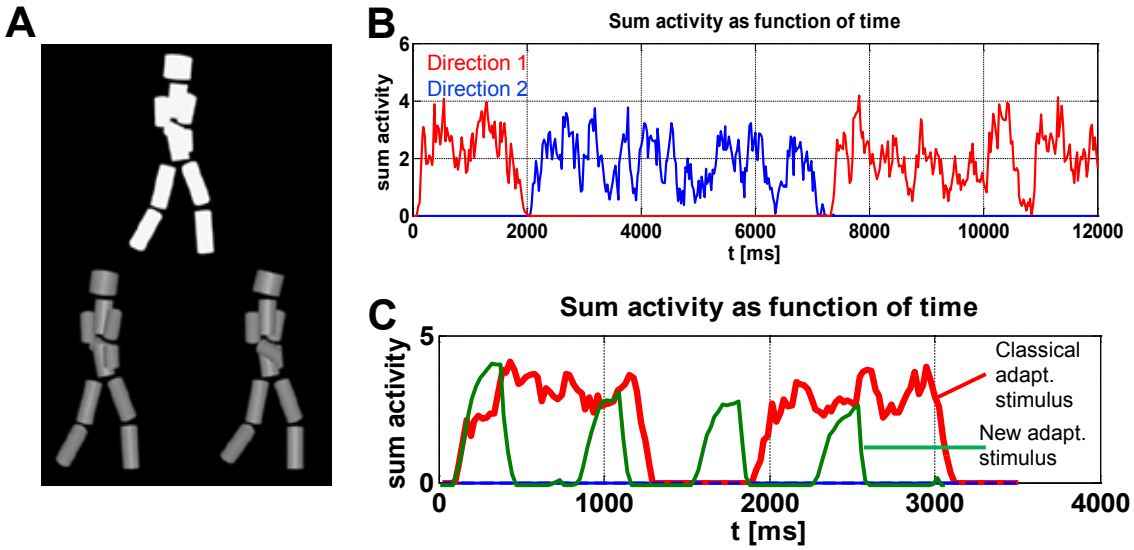


Figure 44: Modeling of the neurodynamics of action perception

A) Ambiguous silhouette stimulus that can be disambiguated by shading cues. B) Perceptual switching reproduced by the neurodynamic model (output activity of two neurons encoding the different views). C) Simulated adaptation effects for action-selective neurons for a normal action stimulus that is repeated once, and for a new type of adaptation stimulus that (according to the model) should, lead to much stronger adaptation effect.

The developed new two-dimensional sequence selective neural field shows a special form of multi-stability that leads to switching between two alternative travelling pulse solutions. We have exploited a new approach to analyse the dynamics of two-dimensional neural field using a level-set approach to analyse this multistable solution (Coombes et al., 2012). This analysis is based on a simplified version of the model above without adaptation and random noise processes. The key idea is a multi-dimensional extension of a method (Amari, 1977) that characterizes the dynamics of localized solutions by their behaviour on the boundary points of the excited region (i.e. the region in the field where neurons are activated). It can be shown that for the two-dimensional case on the boundary contour of the excited region $\mathbf{r}(\sigma)$ (where $0 \le \sigma \le 1$) the neural activity of the field $u(\mathbf{r}(\sigma),t)$ and its gradient $\mathbf{z}(\mathbf{r}(\sigma),t)$ fulfil the differential equation system:

$$\tau_u \dot{u}(\mathbf{r}(\sigma),t) = -u(\mathbf{r}(\sigma),t) + \oint_{\partial B} \mathbf{F}(\mathbf{r}(\sigma)-\mathbf{r}(\sigma'))\,\mathbf{n}(\sigma')\,\mathrm{d}\sigma' + S(\mathbf{r}(\sigma))$$

$$\tau_u \dot{\mathbf{z}}(\mathbf{r}(\sigma),t) = -\mathbf{z}(\mathbf{r}(\sigma),t) + \oint_{\partial B} w(\mathbf{r}(\sigma)-\mathbf{r}(\sigma'))\,\mathbf{n}(\sigma')\,\mathrm{d}\sigma' + \nabla_{\mathbf{r}} S(\mathbf{r}(\sigma))$$

$$\mathbf{z}(\mathbf{r},t) = \nabla_{\mathbf{r}} u(\mathbf{r}(\sigma),t) \qquad \mathbf{n} = \mathbf{z}(\mathbf{r},t) / |\mathbf{z}(\mathbf{r},t)|$$

In this equation the vector field $\mathbf{F}$ fulfils the differential equation $\mathrm{div}\,\mathbf{F}(\mathbf{r}) = w_s(\mathbf{r})$, where $w_s(\mathbf{r})$ specifies the lateral interaction kernel of the neural field for a coordinate system that moves with the traveling solution. By linearization a stability condition can be derived from the last two equations. To our knowledge, this is the first time that such a localized

traveling solution has been analysed within an analytical framework for such two-dimensional field with a relatively general form of the lateral interaction kernel, which in our case needs to fulfil constraints derived from the experimental data, so that not just a mathematically convenient form can be chosen that simplifies analysis. (A manuscript about this work is in preparation; see (Giese et al., 2015a, 2015b) for conference presentations.)

Model for new pathway that accounts for the influence of shading cues on action perception

Shading provides relevant depth cues that influence the 3D perception of human bodies. A known phenomenon in the perception of surfaces is its dependence of light-source direction, and a set of studies has demonstrated that the human visual system embeds a 'lighting-from-above prior' which, in absence of other cues, results in interpretations of shapes that are compatible with illumination from above (D. Brewster, 1847; Ramachandran, 1988). We have discovered a new visual illusion that demonstrates the efficiency of such lighting-from-above priors also in body motion perception. We have developed a new type of body motion stimulus, which is a biological motion stimulus that consists of volumetric elements that can be illuminated from different directions (insets Figure 45A). Consistent with the multi-stability with respect to the perceived view of actions discussed in a), this stimulus without illumination cues results in a bistable percept where the walker is alternately perceive as walking away from the observer into the image plain, or as walking straight towards the observer. An addition of shading cues disambiguates this stimulus. We found that if the stimulus shown in Figure 44A, if illuminated from above, always is perceived with the physically correct veridical walking direction. Changing the light-source direction to lighting from below, however, the walking direction is misperceived as rotated by 180 deg. Lighting from below thus flips the perceived walking direction. Figure 44A shows the accuracy (percent correct classifications) of walking direction as a function of the direction of illumination, which was varied systematically. Further experiments investigated the critical features that drive this perceptual reversal effect, showing that particularly the shading patterns of the thighs and the lower arms are critical. This experiment proves the relevance of internal shading cues, beyond the external silhouette of the body, for the perception of body motion.

We have extended an existing mean-field models for body motion perception (Giese and Poggio, 2003) in order to account for this illusion. Detailed simulation experiments show that the original model, which contains a form pathway that is mainly responding to the outer contours of body stimuli, was not sufficient to account for this influence of the light-source direction. This made the development of a new shading pathway necessary which processes the internal luminance gradients within individual body parts in order to extract 3D shape cues that are helpful for deriving of the walking direction. The neural model for the new postulated pathway is sketched in Figure 45B. It exploits only physiologically plausible computational steps and consists of a hierarchy of filters (deep network). The first level is formed by a layer of simple cells (Gabor filters) that extract contour as well as internal luminance gradients. By a special nonlinear pooling operation a representation of the boundary contours of the body is derived from these cell responses. This contour representation is then used to suppress the influence of the strong luminance gradients on the boundary in the shading pathway by a gating mechanism (since otherwise the pathway's response would be entirely dominated by the luminance edges along the boundary contours of the silhouette). Using the responses of the uneven Gabor filters, a neural population representation of the internal luminance gradients of the moving body is constructed. The resulting filter responses are pooled over neighboring spatial positions in order to make the representation partially position invariant. We apply a feature selection algorithm to the resulting invariant feature vectors, selecting a subset of inputs for radial

basis function units that are trained with the shading patterns of individual keyframes ('snapshots') from training action movies with illumination from above. The highest level of the model is given by a recognition layer with neurons that respond either to walking towards or away from the observer.

The model reproduces the correctly the disambiguation of perceived walking direction if shading cues are present. The model, if trained with stimuli with illumination from above, correctly predicts the perceived walking direction for test stimuli with illumination from above (Figure 45C, upper panel). For testing with walking towards the observer the output neuron in the recognition layer for walking towards (the observer) is much more strongly activated than the one for walking away. If the model is tested with the same stimulus illuminate from below, consistent with the psychophysical results, the perceived walking direction flips, and the walking away neuron is more activated than the walking towards neuron. A detailed analysis of the relevant luminance gradient features allows to understand why this illusion occurs: The gradient features for illumination from above resemble the features for illumination from below for the opposite walking direction. The model also reproduces the results from the experiment investigating the critical feature that drive the visual illusion. Two articles about this work are in writing, and we are presently integrating the new shading pathway in the existing classical architecture (Giese and Poggio, 2003). Poster presentations see, e.g., (Fedorov, 2014; Fedorov and Giese, 2015).

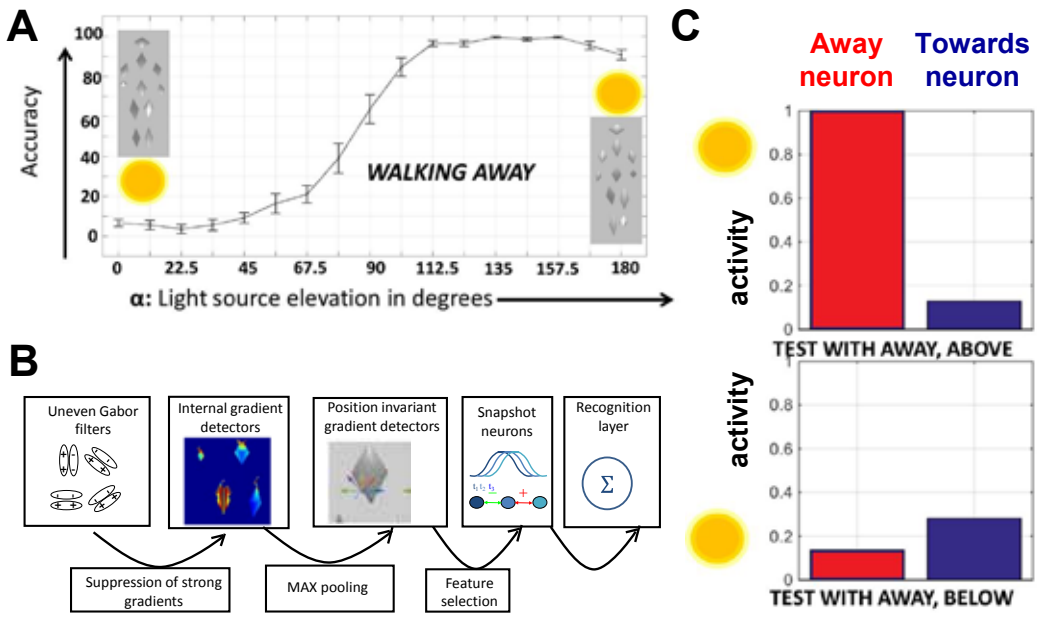


Figure 45: Visual illusion demonstrating lighting-from-above prior in body motion perception and shading pathway for an action recognition model

A) New biological motions stimulus with shaded elements and psychophysical results from a study that investigated the perceived locomotion direction for a walker walking away from the observer in dependence on the light-source direction. Changing the light source from above (180 deg) to below (0 deg) flips the perceived walking direction to walking towards the observer (Accuracy: percent correct direction judgements). B) Model of a new visual pathway for the processing of internal luminance gradients. It consists of a hierarchy of neural detectors, realizing invariance to position by nonlinear pooling. The highest level contains neurons that signal walking towards or away from the observer. C) Simulation result reproducing the illusion. After training with stimuli with illumination from above, the output neurons signaling walking away are more strongly activated by a walking away stimulus than the towards neurons. Testing with the same walking stimulus illuminated from below flips the monotonic order of the output activations, predicting a percept of walking in the opposite direction.

Spiking neuron model for a key circuit linking visual and motor representation of actions

It is well established that visual and motor representations of actions are tightly coupled and potentially even overlapping (Prinz, 1997; Rizzolatti and Luppino, 2001). We have started to develop a quantitative modelling framework for the neural substrate of such interactions within the HBP simulation framework, using the programming package NEST in collaboration with SPs4 and 5 (M. Diesmann and S. Grün). In addition, we have extended experiments that provide data in humans about the influence of motor execution on the visual processing of action stimuli. The data from one key experiment, which investigates the influence of time delays between motor execution and visual input on the detection of action patterns, has been simulated qualitatively in this new neural modelling framework. The efficient implementation of the underlying spiking architecture within the NEST framework is still ongoing, and it is not yet entirely clear if the existing components of this framework permit an optimal implementation of our architecture. Further work in terms possible necessary extensions of the NEST framework are planned. The existing implementations of the model have been provided as part of the HBP model data basis.

The testing of the developed new neural model is based on psychophysical data from an experiment that tested the influence of concurrent motor activity on the detection of point-like stimuli in noise (Christensen et al., 2011). Using a Virtual Reality (VR) setup, participants had to detect a point-light arm in background noise. The movement of the arm was controlled by the subject's own arm movement, which was online motion-captured (Figure 46A). The experiment compared the detection thresholds (in terms of the number of tolerable noise dots) for different conditions that introduced time delays between the movement of the observed arm and the real movement of the participant. In addition, the detection performance without concurrent motor behaviour was tested. Figure 46B shows the results from the experiment, indicating that compared to visual detection without concurrent motor behaviour, a high level of synchrony between motor behaviour and visual stimulus (delay smaller than 300 ms) results in a reduction of the detection threshold while longer time delays, inducing an asynchrony between observed and executed motor behaviour result in an increase of the detection threshold. This dependence of visual detection performance on motor behaviour can be explained by a coupling between visual and motor representations of actions. In more recent work we were able to show a differential involvement of different cerebellar regions in this sensorimotor coupling (Christensen et al., 2014).

As basis for the modelling of such couplings between visual and motor representations in the context of our previous physiologically-inspired models, and as basis for a model which later can be compared to electrophysiological data in monkeys that are presently being prepared at EKUT, we developed a spiking neuron model. This model combines a dynamic neural representation of visually observed actions, consistent with our earlier models (Fleischer et al., 2013), with a dynamic neural representation of motor programs e.g. (Cisek and Kalaska, 2010; Erlhagen and Schoner, 2002). The architecture of the model is based on a neural mass / field model that was implemented by a careful approximation of a mean-field model by spiking networks using biophysically realistic models for single neurons. As model for individual neurons we used an exponential integrate-and-fire model, which was shown before to provide best approximation quality for a benchmark competition, where data from cortical neurons were modelled using different neuron models (Jolivet et al., 2008). The basic unit for the development of the neural mass model are elementary ensembles, consisting of 80 excitatory neurons and 20 inhibitory neurons that are connected randomly, controlling the average connection strength within and between the excitatory and inhibitory populations (Figure 46C). The behaviour of such an ensemble if stimulated with medium input current is illustrated in the right panel of Figure 46C. Many (30) such ensembles are then integrated in a larger network that approximates

a neural field with asymmetric lateral interaction kernel. The kernel determines the average connection strength between the different ensembles dependent on their location within the field (Figure 46D). The form of this lateral interaction kernel and its relationship to the dynamic behaviour of the field has been extensively studied in the mean field limit, where we exploited specifically work on the realization of stimulus-locked traveling pulse solutions in such distributed networks (Xie and Giese, 2002). By extensive simulations, we established that this fully spiking architecture behaves very similar to an Amari type of neural field (by realizing different architectures implementing memory, winner takes all selection, and traveling pulse solutions). For the model we used an interaction kernel that stabilizes a travelling pulse solution, which either is self-stabilizing, or is induced by a travelling external stimulus (for the vision representation). Two fields of this type were then integrated within an architecture that realizes coupled distributed representations of visual and motor patterns (Figure 46E). In the visual field the travelling peak solution follows a travelling input peak that is derived from the visual stimulus information (Giese and Poggio, 2003). In the motor field the travelling solution is induced by a go signals that injects local activity in the field that initiates a self-stabilized autonomously travelling solution. The travelling speeds of both solutions are adjusted in a way so that in the normal case, where the visual input and the motor behaviour are in synchrony, both solutions travel with the same speed. Both fields are reciprocally coupled by interaction kernels that result in a mutual excitation of the fields if the travelling solutions are at the same position along the field, and which induce inhibition if the peak positions strongly differ. As consequence, the motor representation enhances the activity in the visual field when the motor peak propagates with the same speed and phase as the observed visual input.

This simplified model reproduces the results from the experiment described above, as shown in Figure 46F. Compared to baseline (without concurrent motor execution, i.e. without activity in the motor field) higher activity emerges in the visual field when the visual input is synchronous with the motor execution ('synchronous'). However, when the visual input follows the motor execution with a strong delay (560 ms; condition 'asynchronous') the activity in the visual field is substantially reduced. (All differences are significant according to t tests with $p < 0.05$). The model reproduces thus the basic phenomenology of the experiment by (Christensen et al., 2011).

Ongoing work focuses on the following problems: (i) We have started to implement this core model using the HBP simulation software NEST, which is developed as part of the work in SPs 4 and 5. This step was realized in close collaboration with M. Diesmann and S. Grün (Forschungzentrum Jülich) who are members of SP4 and SP5. Meanwhile, individual neuron ensembles have been successfully implemented. However, the software does not optimally support structured connections as required for neural field models, and we are collaborating on efficient ways to implement such networks with speed advantages compared to standard tools like MATLAB. (ii) The demonstrated result is only the first proof of concept for the architecture. We plan to simulate many other experiments and also fMRI results on perception-action coupling with this architecture. In addition, the physiology group in Tübingen will start in the near future recordings that will allow the verification of aspects of the proposed architecture in comparison with single-cell data in monkeys. (iii) Until the end of the Ramp-Up Phase of the HBP we plan to extend the model by a back-end that simulates motor behaviour on a human avatar, and by a front-end that is part of an architecture that we developed previously, and which realizes the recognition of actions from real images and videos (Fleischer et al., 2013).

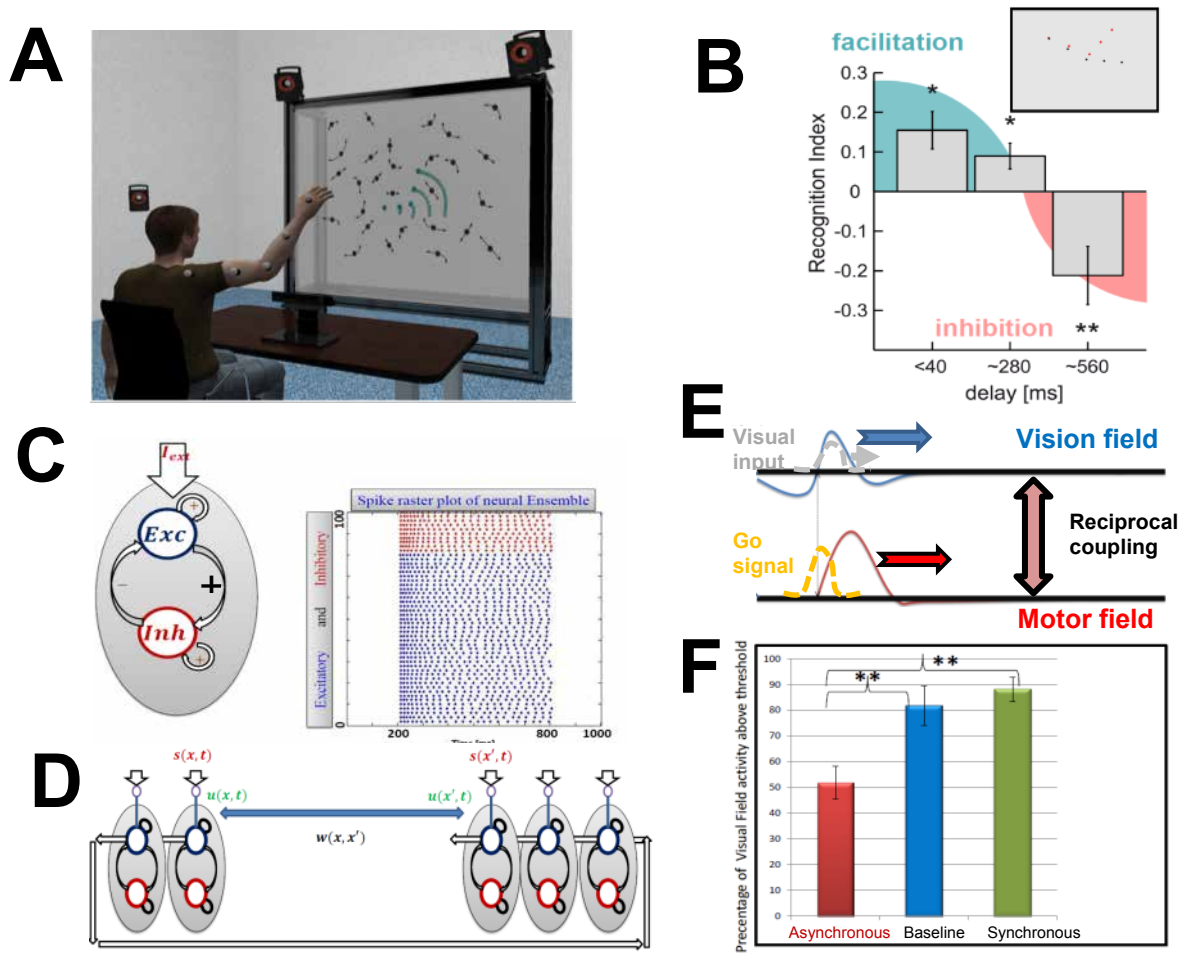


Figure 46: Experiment and spiking neuron model for the interaction between perceptual and motor representations of actions.

A) Experiment investigating the influence of the temporal congruency between motor execution and visual stimulation on the detection of a point-light arm in dynamical noise dots. The stimulus is generated using a VR setup by online motion-capture of the participant's arm movement. B) Main result showing Recognition Index (detection performance relative to baseline without concurrent motor execution) as function of the time delay between visual stimulus and motor execution. C) Ensemble model and generated spike trains consisting of 100 adaptive exponential integrate and fire units. (We could demonstrate similar performance for implementations in MATLAB and NEST environment.) D) Neural field / mass model consisting of 30 coupled ensembles that are mutually coupled by an interaction kernel w(x,x'). E) Architecture consisting of two coupled neural fields, modeling visual and motor representations of actions. The visual field is driven by stimulus activity from body-shape selective neurons. The motor field stabilizes a traveling solution that is initiated by a transient local activity that represents a go signal. Coupling between the fields is mediated by interaction kernels that are adjusted for the modeling of psychophysical data. F) Reproduction of the basic result from the experiment in A. Compared to the condition with silent motor field ('Baseline'), small time delays between visual stimulus and motor execution ('Synchronous') lead to higher activation in the visual representation. In contrast, a strongly delayed activity in the motor representation ('Asynchronous') reduces the amplitude of the traveling solution in the visual representation, and by this the detectability of the visual stimulus.

Data Provenance

The stimulus data basis is available in Tübingen.

The data available to identify key mechanisms for action recognition are available from our own experiments. Electrophysiological data to narrow down mechanisms for the interaction between action execution and recognition are being produced outside the HBP in our own experiments in Tübingen. A proposal to acquire conclusive physiological data about such mechanisms within the HBP SP3 in the future was rejected. Psychophysical data

and neuropsychological that helps to narrow down spatio-temporal tuning of perception-action coupling was provided from our own psychophysical experiments and is partly captured by the developed model. Extension of the model using more data from the literature is in progress and will be continued after leaving SP3.

A Dataset Information Card has been completed (see DIC Task T3.1.2 “Perception Action”).

The data were deposited on a server: http://sp3.s3.data.kit.edu/3.1.2/

1.7 Body Perception and the sense of Self

Task T3.1.3 - Olaf Blanke (EPFL), Nathan Faivre (EPFL), Mel Slater (UB)

Review of the cognitive architecture for bodily self-consciousness

Olaf Blanke, Mel Slater, Andrea Serino "Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness" Neuron, Volume 88, Issue 1, p145–166, 7 October 2015

Abstract

Recent work in human cognitive neuroscience has linked self-consciousness to the processing of multisensory bodily signals (bodily self-consciousness, BSC) in fronto-parietal cortex and more posterior temporo-parietal regions. We highlight the behavioral, neurophysiological, neuroimaging, and computational laws that subtend BSC in humans and non-human primates. This includes body-centered perception (hand, face, trunk), based on the integration of proprioceptive, vestibular, and visual bodily inputs, spatio-temporal mechanisms, and the importance of signal integration within peripersonal space (PPS). We develop four major constraints of BSC (proprioception, body-related visual information, PPS, embodiment) and argue that the fronto-parietal and temporo-parietal processing of trunk-centered multisensory signals in PPS is of particular relevance for theoretical models and simulations of BSC and eventually of self-consciousness.

Data set: Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective

The main goal for this task is to delineate the brain regions associated with the building blocks of bodily self-consciousness, namely self-identification ("owning my body"), self-location (''Where am I in space?''), and the first-person perspective (''From where do I perceive the world?'') (Blanke, 2012). We relied on the full-body illusion (FBI), in which participants feel tactile stimulations on their back, while seeing this stimulation displayed on the back of their own body (a 'virtual body') seen from a third-person perspective (Lenggenhager et al., 2007). In synchronous visuo-tactile stroking conditions, participants typically self-identify more with the seen virtual body, judge their positions as closer to it, and feel the tactile stimulus as coming from it. We used fMRI coupled with robotics in order to measure changes of BOLD signal associated with the FBI. Participants saw a virtual rod moving vertically along the midline of the virtual body's back. A custom-made robotic device generated the same movement profile on the participant's back, either synchronously or asynchronously with the virtual rod (respectively inducing synchronous vs. asynchronous visuotactile stroking). An ultrasonic motor placed at the level of the feet actuated the stimulation sphere over a rack-and-pinion mechanism. Motion was transmitted over a guided fiberglass rod, which held the stimulation sphere over a compliant blade in order to follow the participant's back with constant pressure. In a control condition, the virtual body was replaced by an object, for which no full-body illusion is expected to occur. Therefore, analyses was performed according to a 2 x 2 factorial design, with Object (body; object) and Stroking (synchronous; asynchronous) as main factors.

Differences of activity in synchronous vs asynchronous stroking for body vs. object were found in bilateral temporo-parietal junctions (parietal operculum), right middle-inferior temporal cortex (including the extrastriate body area). In addition, we found an increase of functional connectivity between the bilateral TPJ and the supplementary motor area, ventral premotor cortex, insula, intraparietal sulcus and occipitotemporal cortex (Ionta et al., 2011, 2014). The raw and analysed data from the publication by (Ionta et al., 2011) can be found on our lab server (restricted access), along with Matlab scripts used to perform automatic anatomical characterisation of the activation cluster from (Ionta et al., 2011) and the outcoming results.

In addition to this functional analysis, we performed an anatomical classification of the right and left temporo-parietal junctions (rTPJ and lTPJ). Based on cytoarchitectonically defined regions of the human parietal operculum (OP; (Eickhoff et al., 2006)) and of the inferior parietal cortex (IPC; (Caspers et al., 2006)), the rTPJ cluster overlapped with regions of the OP (OP1: 25.8%) and of the IPC (PFcm: 36.0%, PFop: 16.7%, PF: 8.8%), whereas the rest of the activity was predominantly located on the posterior end of the superior temporal gyrus (pSTG). The lTPJ overlapped with the OP (OP1: 25.8%) and the IPC (PFcm: 21.2%, PFop: 11.9%, PF: 9.6%), again with the rest of the activity in pSTG. OP1, which is presumably the human analogue of area S2 in non-human primates (Eickhoff et al., 2010), is considered to be a "perceptive" area strongly interconnected with the IPC and potentially associated with some of the more complex functions of the OP, such as perceptual learning, tactile working memory and stimulus discrimination. The IPC, on the other hand, is known to integrate basic modalities (somatosensory, visual and auditory) but has also been involved in higher order cognition (e.g. (Silani et al., 2013)). These results were presented at the 2015 HBP summit in Madrid by Nathan Faivre.

Effort is still ongoing to collect and integrate anatomical and functional connectivity data from human and non-human primates to refine our knowledge about the functions of these regions and their contributions to the construction of a coherent image of the 'self'. The data used for these analyses comes from an fMRI study conducted on 22 healthy

participants at the 3T scanner at the University Hospital Center of the canton de Vaud (CHUV), Lausanne, Switzerland. The aim of the study was to assess the neural correlates related fundamental aspects of bodily self-consciousness, namely the subjective self-location in space, the direction of the first-person perspective and self-identification, using a bodily illusion (full-body illusion) relying on multisensory conflicts. The results of this study were published in the Journal Neuron in 2011, along with lesion data from neurological patients suffering from out-of-body experiences. The aim here was to better characterise activation clusters from (Ionta et al., 2011) in terms of brain cytoarchitecture, to be able to refer more specifically to both human and primate scientific literature to find which types of stimuli/experimental paradigms also engage these regions (region-of-interest analysis). The ultimate goal was to try to explain the origin of the relatively complex effects from (Ionta et al., 2011) (subjective self-location and first-person perspective) in terms of processing of simpler bodily signals in the brain. The relevant analyses are completed. All data will be made available to the European Commission together with Dataset Information Cards for the duration of the project and for a period of up to five years after the end of the project by the end of the Ramp-Up Phase.

A **Dataset Information Card** has been completed (see DIC Task T3.1.3 "Neural correlates of self-location and first-person perspective").

Data Provenance

All data were acquired by Silvio IONTA at the *Centre Hospitalier Universitaire Vaudois*, (CHUV – P23) in collaboration with the *Centre d'Imagerie BioMedicale* at EPFL.

Collaborations

We plan to renew a previous collaboration with Wulfram GERSTNER (SP4, EPFL), with whom we previously developed a model of the rubber hand illusion.

Publications

Grivaz, P., Serino, A., & Blanke, O. Meta-analytical assessment of neural correlates of spatial, temporal and social perspective-taking in humans. In prep.

Faivre, N., Doenz, J., Scandola, M., Bernasconi, F., Salomon, R., Bello Ruiz, J., & Blanke O. Illusory Hand Ownership Modulates the Position of After-images: a Case for Self-grounded Vision. Journal of Neuroscience, under review.

Salomon, R., Galli, G., Lukowska, M., Faivre, N., Ruiz, JB., & Blanke, O. (2015). An Invisible Touch: Body-related Multisensory Conflicts Modulate Visual Consciousness. Neuropsychologia, doi: 10.1016/j.neuropsychologia.2015.10.034.

Faivre, N.*, Salomon, R.*, & Blanke, O. (2015). Visual Consciousness and Bodily-Self Consciousness. Current Opinion in Neurology, 28(1), 23-28. (* equal contributors).

Blanke, O., Slater, M. & Serino, A. Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness, in Neuron, vol. 88, num. 1, p. 145-66, 2015.

The Study of Body Ownership and Agency Using Immersive Virtual Reality Methods

Decreased cortical excitability after the illusion of missing part of an arm

Previous studies on body ownership illusions have shown that under certain multimodal conditions, healthy people can experience artificial body-parts as if they were part of their own body, with direct physiological consequences for the real limb that gets 'substituted'. In this study we wanted to assess (a) whether healthy people can experience 'missing' a body-part through illusory ownership of an amputated virtual body, and (b) whether this would cause corticospinal excitability changes in muscles associated with the 'missing' body-part. Forty right-handed participants saw a virtual body from a first person perspective but for half of them the virtual body was missing a part of its right arm. Single pulse transcranial magnetic stimulation was applied before and after the experiment to left and right motor cortices. Motor evoked potentials (MEPs) were recorded from the first dorsal interosseous (FDI) and the extensor digitorum communis (EDC) of each hand. We found that the stronger the illusion of amputation and arm ownership, the more the reduction of MEP amplitudes of the EDC muscle for the contralateral sensorimotor cortex. In contrast, no association was found for the EDC amplitudes in the ipsilateral cortex and for the FDI amplitudes in both contralateral and ipsilateral cortices. Our study provides evidence that a short-term illusory perception of missing a body-part can trigger inhibitory effects on corticospinal pathways and importantly in the absence of any limb deafferentation or disuse.

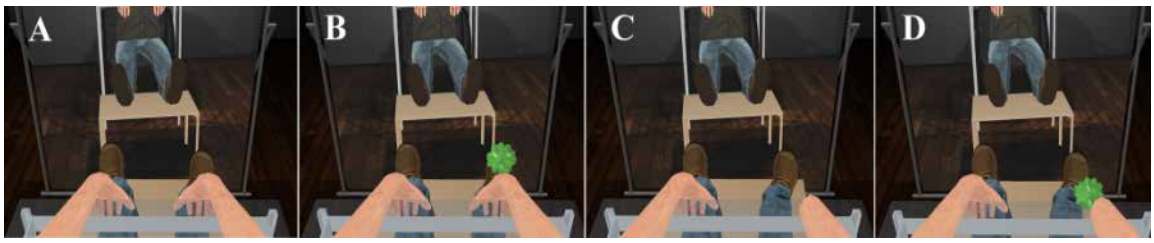


Figure 47

(A) For both groups, a gender-matched virtual body was seen from 1PP with the same posture with the real body, as if they were spatially coincident. (B) For both groups, a virtual ball touched the right virtual hand various times while the real hand was physically touched at the same timings (2 minutes). (C) For the amputation group, part of the right virtual arm disappeared, as if amputated (3 minutes). (D) Following, the virtual ball touched the area previously occupied by the right virtual hand and part of the forearm (i.e. the table surface) without triggering any physical touch to the participant's right hand (10 minutes).

The session started for both groups identically. Through the HMD, participants saw a complete gender-matched virtual body from a 1PP, spatially coincident with their real body. In order to make the participants' posture as comfortable as possible while being motionless, a virtual mirror was placed just opposite so that they could see the virtual body both when looking down but also when looking straight ahead (Figure 47). Participants were asked to move their head and describe what they saw around themselves, including the virtual body (1 minute). Once the scene was described, the experimenter asked them to focus on the virtual hands for the rest of the session either by looking either directly or through the mirror. Then, a virtual ball appeared and tapped various times the right virtual carpal, bouncing between the virtual hand and the virtual mirror in random velocities for 2 minutes (Figure 47). Every time the ball made contact with the virtual carpal and for the whole duration of this contact, the vibrator on the participants' right hand was triggered and the *vibrationSkin* sound file was reproduced. Therefore, all seen and heard tactile events on the virtual hand were temporally registered with physical touch on the real hand.

Overall questionnaire scores showed that participants in the condition where the virtual arm was missing tended to agree with statements supporting this. An *amputation* score, calculated as the average of the five amputation statements, was significantly and positively correlated with arm ownership (*ownarm* - rs=0.528, n=40, p<0.001) and full body ownership (*ownbody* - rs=0.464, n=40, p=0.002)), and with feelings of not being able to move the right hand (*nomoveright* - rs=0.689, n=40, p<0.0001). No other significant correlations were detected between the amputation score and the rest of the items. See Figure 48.

For the EDC muscle, there was no main effect of condition ($F_{(1,37)}$=0.45, p=0.507), nor of time ($F_{(1,37)}$=0.56, p=0.460) but a main effect of hemisphere ($F_{(1,37)}$=4.76, p=0.035) was detected. None of the two or three way interactions were significant. Residuals errors were not normally distributed (Shapiro Wilk test, p<0.0001). Visual inspection of the residuals' plot clearly identified four outlier values. After removal of the outliers, the ANOVA revealed again no main effect of condition ($F_{(1,36)=}$0.58, p=0.452), nor of time ($F_{(1,36)=}$=0.11, p=0.747) but a main effect of hemisphere ($F_{(1,36)=}$=4.68, p=0.037, partial η^2=0.370) was detected. None of the two-way interactions were significant. However, the three-way interaction condition × time × hemisphere had significance level p = 0.053 ($F_{(1,35)=}$=3.98). Residual errors were normally distributed (Shapiro Wilk test, p=0.575). Post-hoc t-tests revealed that the EDC amplitudes for the left hemisphere (right hand) were significantly lower than those for the right hemisphere ($t_{(74)}$= -4.61, p< 0.001, CI[0.13, 0.33]).

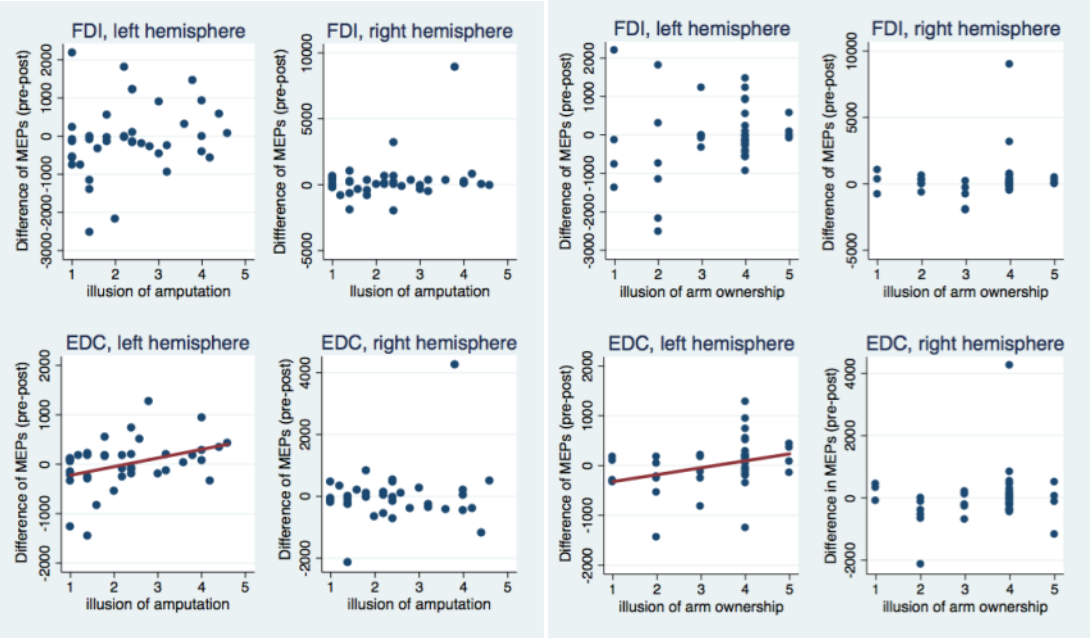


Figure 48: Scatterplot of the difference between and pre and post-VR MEP amplitudes with the illusion of amputation per each muscle and hemisphere (on the right).

The difference in amplitudes was significantly correlated with the amputation illusion only for the EDC muscle of the left hemisphere. Similarly, only in this case, the amputation illusion predicted significantly the MEPs' difference (b = 174.33, t =2.53, p = 0.016) and explained a significant proportion of its variance (R^2 = 0.12, F(1, 37) = 6.39, p = 0.016). Scatterplot of the difference between and pre and post-VR MEP amplitudes with the illusion of arm ownership per each muscle and hemisphere (on the left). The difference in amplitudes was significantly correlated with the arm ownership illusion only for the EDC muscle of the left hemisphere. Similarly, only in this case, the ownership illusion predicted significantly the MEPs' difference (b = 139.27, t =2.10, p = 0.043) and explained a significant proportion of its variance (R^2 = 0.08, F(1, 37) = 4.40, p = 0.043).

Our study demonstrates the possibility of inhibitory effects on corticospinal pathways, triggered by a short-term illusory perception of missing a body-part and importantly in the absence of any physical intervention (e.g. ischaemic nerve block or movement restriction); just a few minutes of a novel body experience seem to be sufficient to produce direction specific effects on the excitability of sensorimotor system. Although future studies are needed to investigate whether the origin of the effects is primarily cortical or spinal, as well as their temporal prevalence, our results indicate new possibilities for inducing short-term functional reorganization and plasticity of the sensorimotor system by emphasizing the contribution of body experience. Experience is a known modulator of the brain in both functional and structural terms (Makin et al., 2013; May, 2011) and its role has widely discussed in relation to increased use, disuse and transient deafferentation of a body-part in healthy participants . A novel perception of the body that suggests that a body-part is missing may also entail psychologically both the body- part's disuse (inability to move it) and deafferentation (inability to receive afferent input from it), since it is no longer there

Illusory Agency over Walking

Figure 49: Illusory Walking setup

(A) Participants were seated still on a stool, except for head movements. (B) Initially participants saw the standing virtual body reflected in a mirror (1PP condition). (C) Participants saw the walking virtual body from 1PP, or (D) from 3PP. The walking body always cast a shadow.

Here we show that participants can have the illusion of agency over the walking action of a virtual body even though in reality they are seated and only allowed head movements. This paper follows on from previous work in (Banakou and Slater, 2014). This between-groups experiment had two binary factors (Figure 49): Perspective (1PP or 3PP) and Head Sway (Sway or NoSway). In 1PP participants saw a life-sized virtual body spatially coincident with their real body from a first person perspective. In 3PP participants saw the virtual body from third person perspective (3PP). In the Sway condition the viewpoint followed a walking animation, and in NoSway it was solely determined by head movements. The results show strong illusions of body-ownership, agency and walking, in the 1PP compared

to the 3PP condition. Sway reduced the level of agency. In order to test the effects of perspective and body-ownership on the feeling of agency over walking movements, we immersed participants in a virtual environment, where a virtual body, seen either as spatially coincident with the real body and from a first person perspective (1PP) or separate from the viewpoint of the real body from third person perspective (3PP), was walking forward across a field (Figure 49). Moreover, in order to test the importance of the optic flow, a second factor was whether a sway animation was applied or not to their viewpoint. This sway was based on a pre-recorded animation of real walking. In other words there was a sway applied to the head as one factor (Head Sway) that had two levels Sway or NoSway.

A questionnaire on body ownership and agency resulted in strong illusions of ownership over the virtual body, and agency over the walking in the 1PP but not in the 3PP conditions. The Sway factor reduced the illusion of agency.

Participants experienced the walking for 4 minutes. For the first three minutes the walking was over level ground, but then continued up a hill for 44s. We predicted that agency over the walking would result in heightened physiological responses during the hill climbing period compared to a baseline period that started 90s before the hill climbing and lasted for as long (44s). We recorded skin conductance, ECG and respiration. Physiological data from one participant (belonging to the Sway condition) were not available due to a failure in recording. To compare the responses across the conditions, we used as a response variable the differences between the mean skin conductance amplitudes in the hill-climbing period and the baseline (*dSC = mean(Hill Climbing) - mean(Baseline)*) (see Methods - Response Variables). A mixed effects ANOVA for *dSC* on Perspective and Sway and their interaction found no effect for the interaction or Sway terms (P = 0.87 and 0.96, respectively) but for Perspective P = 0.025. The result for Sway does not change when the interaction is removed, and Perspective has P = 0.005 when Sway is also removed. The coefficient for Perspective (3PP=0, 1PP=1) is 0.28 ± 0.10 (SE), with 95% confidence interval 0.08 to 0.48. This reflects what can be seen in Table 1: that 1PP resulted in greater change in SC than 3PP, but that there are no other effects.

	NoSway	Sway	Overall Mean
3PP	0.40 ± 0.31	0.17 ± 0.15	0.29 ± 0.17
1PP	0.64 ± 0.27	0.64 ± 0.24	0.64 ± 0.17
Overall Mean	0.52 ± 0.20	0.41 ± 0.15	0.44 ± 0.09

Table 1: Means and SEs of change in skin conductance (microsiemens) from baseline to hill climbing (dSC).

Our results show that even while participants are seated in a chair and not walking, seeing the virtual body from a first person perspective when it is walking can result in high levels of body-ownership and self-attribution of the walking action. This is supported by both subjective and physiological responses.

Our results highlights an important difference from the setup of Banakou and Slater [3] since

from the outset participants saw their virtual body doing something that they were definitely not doing, whereas in the earlier study the talking occurred only after several minutes into the experiment. Agency only on the basis of 1PP is unlikely. This might be explained by the work of Patla and colleagues who explored the importance of vision of the body while walking (Patla, 1997; Patla et al., 1996). According to their results, viewing the limb position and movement plays an important role in planning and regulating the swing limb trajectory. Given this notion, it is possible that when the legs of the collocated virtual body (1PP) were observed while walking forward, an action representation for the planning of the next movement might have been initiated. Actually, this is not surprising at all, since in our whole life when we look down and see our legs walking, we are walking. Hence, it is possible that a combination of the seeing the walking legs plus possible intention created by the walking experience contributed to the illusory agency. Further studies will need to carefully investigate this possibility and the relative importance of the two.

The work on agency is continuing in collaboration between EPFL (Dr Olaf Blanke's group – Task T3.1.3) and UB as our final experiment in the HBP. We are preparing an fMRI study, based on the experiment of (Banakou and Slater, 2014) in order to assess brain correlates of this type of illusory agency.

The questionnaire and physiological data from these experiments will be publicly available on the web sites of the open access journals to which the papers have been submitted (assuming that they are eventually accepted).

The following paper has been accepted now:

Kilteni K, Maselli A, Kording KP and Slater M (2015) Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Front. Hum. Neurosci*. **9**:141. doi: 10.3389/fnhum.2015.00141

When the above paper is finally published, the data set will be deposited as 'supplementary information' with the paper. The other paper is still in review in Scientific Reports, and if it is accepted then the data will be available there.

Motivation, Decision and Reward

WP3.2 Coordinated by Mariano Sigman

This work-package aims at making progress in understanding how decisions are computed, and how these computations are implemented in neural circuits.
A common computational approach subtends the four tasks of this work-package. All use the computational level as a starting point to model decisions. Models are used to formalize decision algorithms, and theoretical predictions from different models are compared with actual behavioral data to identify the best computational models, and also the best fitting values to parameterize these models. These models can then provide quantitative estimations of the variable computed by the decision algorithms, which in turn are used to investigate their neural underpinnings.
The different tasks also share a common conceptual framework. Decision algorithms can be cast as a maximization problem: they aim at maximizing accuracy, which in many situations also corresponds to maximizing reward associated with correct decisions.
Motivation is the process that translates a reward prospect into the invigoration of the processes involved to obtain this reward. Motivation therefore plays a key role in decision making, both in normal and impaired condition. The group of Talma Hendler (T3.2.4) investigates this process at different scales in humans, from neuronal spikes recorded with implanted electrodes, to macroscopic neural assemblies recorded with surface electrophysiology and functional magnetic resonance imaging (fMRI). The group of Mathias Pessiglione (T3.2.2) investigates motivation from the clinical viewpoint, by characterizing patients with different motivational disorders.
Maximizing accuracy and improving decisions is also typically achieved by learning and taking advantage of the data at hand at any given moment. This requires the detection of one's own error to improve future decisions. Importantly, the likelihood that a given decision or estimate is erroneous can also be anticipated, prior to any feedback. Task T3.2.1 investigates how confidence in one's own computation is estimated and used in a variety of contexts. Florent Meyniel, Stanislas Dehaene and Mariano Sigman investigate confidence in perceptual decision and in statistical learning in humans using fMRI; the group of Rui Costa investigates confidence in motor task performed by rodents. In task T3.2.3, the group of Tobias Donner investigates how the brain weights the incoming, momentary evidence to make perceptual decisions. They explore the possibility that such gating of information relies, at least in part, on the modulation of brain-scale cortical networks by neuromodulators such as noradrenaline, whose endogenous release can be tracked non-invasively with pupillometry and brainstem fMRI in human.

2.1 Mapping and understanding the neuronal circuits involved in decision making, confidence and error correction

Task T3.2.1 - Mariano Sigma (CEA), Florent Meyniel (CEA), Rui Costa (FCHAMP), Rodrigo Freire Oliveira (FCHAMP)

Review of the cognitive architecture for decision and confidence

Florent Meyniel, Mariano Sigman, Zachary F. Mainen "Confidence as Bayesian Probability: From Neural Origins to Behavior", Neuron, Volume 88, Issue 1, p78–92, 7 October 2015

Abstract

Research on confidence spreads across several sub-fields of psychology and neuroscience. Here, we explore how a definition of confidence as Bayesian probability can unify these viewpoints. This computational view entails that there are distinct forms in which confidence is represented and used in the brain, including distributional confidence, pertaining to neural representations of probability distributions, and summary confidence, pertaining to scalar summaries of those distributions. Summary confidence is, normatively, derived or ''read out'' from distributional confidence. Neural implementations of readout will trade off optimality versus flexibility of routing across brain systems, allowing confidence to serve diverse cognitive functions.

Data set 1: Confidence during probabilistic reasoning, behavioural and fMRI recordings

Rational of the work

The sense of confidence has been defined as "a belief about the validity of our own thoughts, knowledge or performance that relies on a subjective feeling" (Grimaldi et al., 2015). It is the capability of the brain to estimate the reliability of its own processing, and use such estimation to optimize further processing and behaviour. We reviewed the current state-of-the art and proposed a general conceptual framework, or "cognitive architecture", to understand the implementation of the sense of confidence in the brain (see above; (Meyniel et al., 2015a)). Here we summarize the main points that motivated the three experimental studies reported below.

Study #1 Choice and confidence readouts in a perceptual decision. Our review stressed that a principled origin for confidence information in perceptual decisions should be the very sensory data that guide decisions. We therefore postulate that there should be two different readouts of the same data: one to select a response (readout of choice) and another one to quantify the level of evidence supporting this decision (readout of confidence). This view is parsimonious, but is poses a conundrum: if choice and confidence are readouts of the same sensory data, why does confidence sometimes depart from choice accuracy? We suggest that reading out choice and confidence from sensory data is unlikely to be innate, and instead it should be learned, which leaves room for imperfections and biases (Baranski and Petrusic, 1994; Meyniel et al., 2015a). In this study, we specifically aimed to test another source of dissociation between confidence and choice: different aspects of the same data may be processed (read out) differently. We capitalized on a previous study in which sensory data at each trial was provided as a series of data samples. Behavioural results suggested that subjective confidence level stemmed mostly on initial data samples, whereas choice stemmed on a more protracted integration of data samples (Zylberberg et al., 2012). Here, we conducted a new study to test this effect while recording brain activity with fMRI. Note that we selected a perceptual decision paradigm, since our review stresses that such paradigms are well suited to test humans and non-humans animals (Kepecs and Mainen, 2012a; Kepecs et al., 2008; Kiani and Shadlen, 2009a).

Study #2 A normative account subjective confidence during probabilistic learning. Human subjects (and maybe other animals) experience confidence feelings when they are engaged in perceptual decisions; however, such feelings extend to other cognitive process, although they receive little attention in neuroscience. Confidence in the course of learning is an illustrative example: this subjective feeling has been overlooked both in the traditional fields of confidence studies, but also in learning. We adopted Bayesian computations as our conceptual framework and we hypothesized that learning some quantity and estimating whether this inference is accurate should both derive from the same algorithm. In other words, they should be two different readouts of the same inferential data. Note that this proposal is similar to the one made above in the perceptual domain, but it is extended to statistical learning. We designed a new behavioural task and the optimal solution of this problem – the so-called Bayesian ideal observer – to test the normative origin of confidence feelings in learning and characterize their properties.

Study #3 A computational role for confidence in the brain during probabilistic learning. While studies #1 and #2 investigate how confidence is estimated by humans subjects, Study #3 investigate how confidence is used, an aspect that is often overlooked (Meyniel et al., 2015a). We followed-up study #2 to probe the computational role of confidence in learning, and its functional consequences in the brain. In any learning

situation, the new incoming evidence must be balanced against prior knowledge in order to update what has been learned. This weighting of information is critical in a world that is both stochastic and changing, where observations are governed by probabilities that change over time. If one favours the incoming evidence too much, the learned estimates will be dominated by random fluctuations in momentary evidence instead of converging to the true underlying probabilities. Conversely, excessive reliance on the previously acquired knowledge will slow the learning process and impede a quick reset when the environment changes. A normative solution to this general combination problem requires weighting each source of information according to its reliability (Jaynes, 2003; Knill and Pouget, 2004; Ma et al., 2006; Meyniel et al., 2015b; Pearl, 1988). In the context of learning, we propose that the subjective sense of confidence in what has been learned serves in the brain as a weighting factor to optimally balance prior and incoming evidence.

Experimental and theoretical work

Study #1 Choice and confidence readouts in a perceptual decision (behaviour and fMRI)

We first adapted the seminal luminance discrimination task by (Zylberberg et al., 2012) to fMRI (Figure 50A). Two pilot subjects were run. The psychophysics results were consistent between subjects and similar to what is shown in Figure 50B. However, we could not find any correlate of perceptual evidence in their visual cortex. We reasoned that since the early visual cortex encodes contrast rather than luminance (Hubel and Wiesel, 1962), our perceptual task may not drive sufficient activity in fMRI. We therefore modified the task and used Gabor patches with slightly different contrast levels (Figure 50A). Six subjects performed this version of the task in the fMRI. We also run a localizer to identify, for each subject, the portions of their visual cortex that responded retinotopically to the location of Gabor patches. During the task however, the difference in contrast levels between the left and right Gabor patches did not elicit a noticeable lateralization of fMRI signals in the visual cortex, even when focusing specifically on the regions defined with the localizer. Given that behavioural performance was kept at a 75% of correct trials by adapting the difficulty for each subject, our negative fMRI result suggests that the corresponding difficulty level elicits differences that cannot be detected with the current signal-to-noise ratio of fMRI (despite our 1.5mm isotropic resolution, and a sampling frequency of 0.5Hz). Since tracking a neural variable corresponding to perceptual evidence could not be achieved, we aborted the study. We nevertheless report below the behavioural results. Subjects' choices were parametrically impacted by the direction of the evidence presented: whether the right patch is more contrasted than the left one on average in a given trial. Confidence, on the other hand, was impacted by the strength of the evidence supporting the choice, be it “right” or “left”: the absolute average difference between contrast levels in a given trial. We also performed a similar analysis for each sample of evidence presented (instead of their average per trial) to characterize the so-called choice and confidence “kernels”. Interestingly, each sample contributed fairly equally to choices, whereas confidence was impacted more by the first samples presented (Figure 50B, bottom). These results concur with the previous data by (Zylberberg et al., 2012): choice and confidence seem to derive from the same sensory data, but with dissociable readouts.

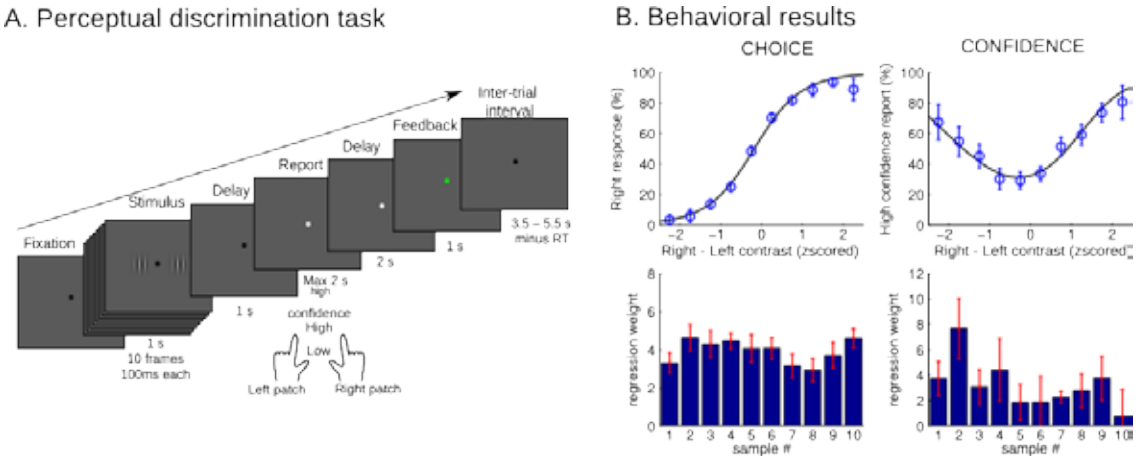


Figure 50: Choice and confidence in a perceptual task with multiple samples

(A) Task design: at each trial, subjects discriminate whether the Gabor patch presented on the left or right is on average more contrasted, knowing that ten samples are provided. In a given trial, contrast levels are sampled from two Gaussian distributions with equal variance and slightly different means. Subject reported at the same choice (left/right) and confidence level (high/low) with dedicated response buttons. (B) Psychophysics results for choices (left) and confidence reports. The top row shows choice and confidence as a function of the average difference in contrast level at each trial. The bottom row shows regression weights of a logistic regressions (linear effect of contrast difference for choice; quadratic effect of contrast difference for confidence) computed independently for each sample position. The plots show mean and s.e.m.

Study #2 A normative account of subjective confidence during probabilistic learning (behaviour)

To investigate confidence in learning, we developed a new statistical learning task, in which subjects (n=18) occasionally reported (1) statistics that their learned from the input they were presented with and (2) the subjective confidence level associated with their estimates (Figure 51A). The input was a sequence of binary stimuli and critically, the statistics of the input changed over time. This instability of the statistics induced variations of confidence levels over time. Our working hypothesis was that learning and estimating confidence in what has been learned both derive from the same inference. We used as a benchmark inference the optimal Bayesian inference (the so-called Ideal Observer), which computes with full probability distributions. Confidence in this framework can be formalized as the log precision of the posterior probability distribution.

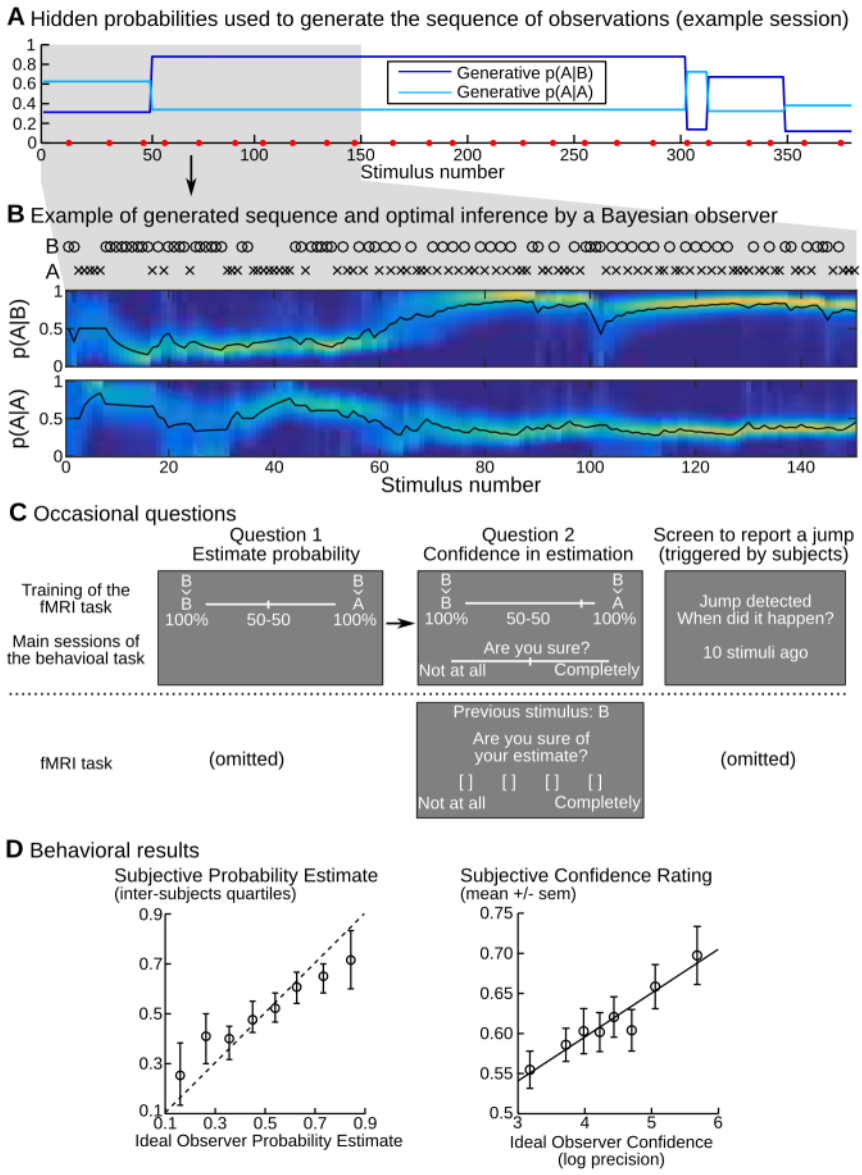


Figure 51: Confidence in a probabilistic learning task

(A) Sequences of two stimuli (As and Bs) were generated randomly following specific transition probabilities. These probabilities were selected randomly, and they were fixed only for a limited period of time: they changed unpredictably. (B) Example sequence of stimuli generated. The transition probabilities that are learned from this sequence by an Ideal Observer are presented as likelihood distributions, as a function of the number of stimuli observed so far. (C). Questions were occasionally asked to subjects (ex. at trials indicated by red dots in panel A) about their estimates of transition probabilities and their confidence in those estimates. Subjects also reported when they detected changes in the probabilities. The fMRI version of the task (bottom) was a simplified version of the original behavioural task (top): subjects were only asked about their confidence in their probability estimates, although these probability estimates remained covert (they were not reported to the experimenter). (D) Behavioural results of the "behavioural task", adapted from (Meyniel et al., 2015a).

The results were published (Meyniel et al., 2015b); the main findings are outlined below. Subjects reports related linearly to the optimal statistical estimates, and also to the optimal confidence levels in these estimates (Figure 51D). Subjects also reported changes in the statistics that they tracked based on the evidence conveyed by the sequence of observations and quantified with the Ideal Observer. Subjective confidence reports fulfilled several properties of an optimal inference: (1) confidence increased with the amount of observations received since the last change (2) confidence decreased when

optimal learned estimates of the statistic changed from one observation to the next and (3) confidence was lower when the statistics made observations less predictable (it is technically the effect of entropy). Interestingly, subjective confidence still covaried with optimal confidence levels even when the three above effects were linearly regressed out, suggesting that subjective confidence is more sophisticated than this list of factors taken together that instead, it may derive from the full probabilistic inference itself. Last, the accuracy (with respect to optimal levels) of statistical estimates and confidence levels were positively correlated. This was true across subjects and also across trials of each subject, indicating that both statistical estimates and their associated confidence levels share a common origin. Overall, our data support the hypothesis that confidence in learning is derived from the learning algorithm itself, and that this algorithm is fundamentally probabilistic and that it operates close to the optimum.

Study #3 A computational for confidence in the brain during probabilistic learning (fMRI)

In this last study, 21 subjects performed the task presented in Figure 52 in an fMRI scanner, in order to probe the functional consequences of different confidence levels in the brain during learning. This work was presented in conference talks and submitted for publication, we summarize here the main findings.

Our central computational assumption is that confidence in what has been learned serves as a weighting factor in the updating process, to balance our prior knowledge on the one hand, and the momentary incoming evidence on the other. One should update more the current estimates when new observations are highly surprising – this is the logic of any learning algorithm (Rescorla and Wagner, 1972; Sutton and Barto, 1998) – and when there is low confidence in the current state of knowledge – which is sometimes referred to as adjustable learning rates (Sutton, 1992), a feature that is handled automatically when one computes with full distributions (Meyniel et al., 2015a) as illustrated in Figure 52A. Regions whose activity co-vary with the updating process were identified by regression analysis using the optimal levels of update computed by the Ideal Observe. A fronto-parietal network was identified, see Figure 52B. We then refined our description of this process by identifying regions whose activity tracked more specifically the likelihood of current observations (the “surprise” levels) or the uncertainty in the current knowledge (the “confidence” levels) or both. We sorted trials according to the Ideal Observer estimates so as to reveal unique signatures of a “confidence” signal, a “surprise” signal and an “update” signal, see Figure 52C, and use these expected profiles as a diagnostic tool to characterize brain regions. Several regions tracked “surprise” (the SMA, FEF, pSTS), other tracked “confidence” (the IPS, DLPFC, OFC) and other the combination of surprise and confidence (IFG). A follow-up analysis of the IFG revealed that its activity may be subtended by a hierarchical inference. We can test unambiguously this aspect in our task because the statistics that must be learned were transition probabilities, and not simple frequencies, which makes specific predictions in the case of hierarchical inference. Together, our results suggest that a sophisticated probabilistic learning algorithm may indeed be implemented in the brain.

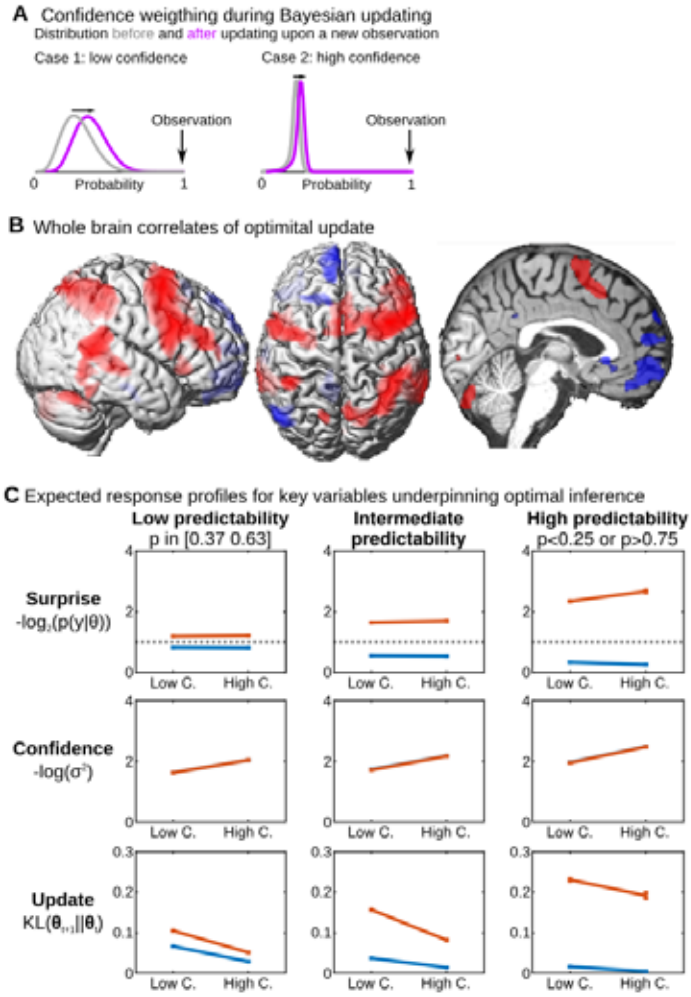


Figure 52: Confidence weighting of surprise during learning

(A) Update of the estimated probability distribution upon observing a new outcome. Note that the shift before (grey) / after (purple) is larger when confidence in the initial estimation is lower (case 1 vs. 2). (B) fMRI signal correlating positively (red) and negatively (blue) with the optimal level of update computed by the Ideal Observer during the probabilistic learning task. (C). While surprise, confidence and update are all correlated, focusing on particular trials reveals unique expected responses for each of these variables. Trials were sorted based on three factors: predictability of the outcome (low, medium, high), confidence in the probability of the outcome (low, high) and expectation violation (the outcome was expected – blue; or unexpected – orange). The values plotted here were computed with the Ideal Observer across all sequences and stimuli presented to our subjects.

Research outputs

Here are the main conclusions of our studies:

- Choice and confidence seem to derive from the same sensory data and with dissociable readouts.
- This hypothesis of a common origin seems general and extends in a non-perceptual domain like abstract statistical learning: confidence in learning may derive from the same inferential data as the single point estimates reported by subjects.
- We can provide a normative account of confidence in learning. Several learning algorithms that work with single point estimates only cannot account for learning in humans. Instead, the human learning algorithm seems essentially probabilistic – this feature should be explored more in the future.
- Confidence in learning is not only estimated accurately, it also seems to play a specific computational role in the learning algorithm: it balances prior and current evidence.
- Functional correlates of confidence in the brain suggest that confidence modulates the weights of previous estimations and the momentary likelihood of observations in the updating process.

These results were disseminated as follow:

- Study #1 (confidence in learning): given that fMRI is not an appropriate tool to track perceptual evidence at the neural level in this kind of task, Mariano Sigman and Florent Meyniel extended their project and invited Tobias Donner (University of Hamburg, Germany; also member of this work package) to follow-up this collaboration and test whether MEG would be better suited to track perceptual evidence in the brain. A noticeable advantage of MEG over fMRI is that its temporal resolution may allow to track the contribution of each individual sample of evidence that are presented sequentially to subjects.
- Study #2 was published in Plos Computational Biology by F. Meyniel, D. Schlunegger and S. Dehaene. It presented by F. Meyniel at a joint SP3-SP4 workshop (EITN, France June 2014). Study #2 and study #3 both suggest that learning in the brain relies on a powerful probabilistic algorithm. This feature motivated the organization of a joint, two-day SP3-SP4 workshop (organized by S. Dehaene, F. Meyniel (SP3) and A. Destexhe, W. Maass (SP4 – EITN) at Collège de France, France, September 2015) with several international speakers to share current results and theories about probabilistic inference in the brain. Both studies were also presented by F. Meyniel at a workshop on confidence organized by S. Dehaene and Z. Mainen (Les Treilles, France, June 2015) and at a workshop on confidence at the Cosyne conference (Salt Lake City, USA, February 2016). Study #3 has been submitted for publication (Meyniel and Dehaene).

A **Dataset Information Card** has been completed (see DIC Task T3.2.1 “Human networks involved in confidence (fMRI and behaviour)”).

Data Provenance

The fMRI and behavioural data were collected by Florent Meyniel at Neurospin, CEA, France.

Location of our data storage

The data were made available:

- Study #1: data were deposited on a server at http://s3.data.kit.edu/SP3/3_2_1/Study1_PerceptualConfidence
- Study #2: data are available as on-line supplementary material of the Plos

Computional Biology publication; they were also deposited on a server at http://s3.data.kit.edu/SP3/3_2_1/Study2_ProbabilitistLearning_behavior

- Study #3: data were deposited on a server at http://s3.data.kit.edu/SP3/3_2_1/Study3_ProbabilisticLearning_fMRI

Data set 2: Confidence estimation on motor skill performance in mice

Abstract

When a musician performs a difficult piece, actions have to be performed with striking accuracy. Previous work has shown that after a self-paced action sequence has been rehearsed, the trial-to-trial variability in performance decreases (precision increases). Furthermore, if the action is complex the modulation of behaviour and neural variability is contingent to the relevance of each dimension (e.g. speed, duration) to the task. However, it is unclear whether animals can monitor and report their performance on a particular trial before the outcome is presented or not. We train mice to execute 4 or 5 sequential presses in order to obtain a cached reinforcement. After training, animals are asked to wait for 8 secs before the outcome is known; animals can wait to know the outcome or abort the trial and start again. Mice abort more trials after incorrect than correct sequences. Logistic regression analysis shows that the probability of current trial abortion depends on the recent history of trial abortion (with slow dynamics) and current trial performance (with faster dynamics). These results show that mice learn to perform sequences of movements within narrow constraints and that they are capable of monitoring their own performance in the absence of outcome. Moreover, the data shows that variables with different time dynamics are involved in assessing action performance.

Introduction

In various daily endeavours (e.g. sports and music) humans perform uniquely accurate actions so that an expected outcome can be achieved. These complex actions require extensive training before an expertise level is reached. Even after such actions are mastered they are seldom performed fully automatically, on the contrary, continuous monitoring allows for compensating unpredictable noise (either in the environment, sensory perception or in the motor commands) (Miall and Wolpert, 1996; Maidhof, 2013; Maidhof et al., 2013; Ma and Jazayeri, 2014). In addition, when the effort/cost of such actions is high, monitoring helps in informing how likely the current performance is to fall within the acceptable range which generates the expected outcome (Giovanni Pezzulo, 2012) and promotes a stronger sense of agency (Demanet et al., 2013). Monitoring can be further helpful in allocating the energy/vigour when facing similarly demanding tasks with similar internal states (e.g. motivation, deprivation or attention) (Krebs et al., 2012; Salamone and Correa, 2012; Skvortsova et al., 2014; Varazzani et al., 2015; Verguts et al., 2015). It is unknown if the capability to learn and execute this class of motor skills is uniquely human. Nor it is known the dynamics of the variables modulating the monitoring of action policies or what are the brain areas responsible for keeping track of one's own actions and calculating confidence on one's own performance.

So far, confidence estimation has been studied in perceptual decision making tasks including its neural substrates in primates (Kiani and Shadlen, 2009b); in rodents (Mainen and Kepecs, 2009; Lak et al., 2014) and; in humans (Hebart et al., 2014; Meyniel et al., 2015b). The estimation of confidence after one’s own performance has received considerably less attention and has, to our knowledge, not yet been studied in rodents. In order to properly investigate the neural substrates of confidence estimation of actions a robust behavioural rodent assay for confidence estimation in action performance was developed. In this operant assay mice perform self-paced trials where, first, an action sequence is executed and later the confidence about the motor skill execution is explicitly reported.

In this manuscript, we describe the task, results, analysis and a computational model that helps explaining the temporal dynamics of the variables mediating confidence computation.

Results

Accurate motor skill learning in mice

Animals (N = 12, BL6/C57) were introduced to the operant box (Med Assoc.) for 30 min (magazine training; MT, Figure 53A) where reinforcements (10% sucrose solution) were delivered in a pseudo random schedule. Next, animals were trained in continuous reinforcement (CRF) for 5 days, where each lever press was reinforced. Next, for 13 days animals were trained in an accurate fixed ratio schedule (AFR45) where only sequences of either 4 or 5 consecutive presses uninterrupted by visits to the magazine lead to the delivery of the cached reinforcement when animals reached the magazine (recorded by an IR beam break). The target (4 or 5 consecutive presses) was covert and animals had to explore different sequence combinations until the covert target was finally learned. Incorrect sequences resulted in no reinforcement delivered and the sequence was reset. Furthermore, the sequences were self-initiated and self-paced. Sessions were finished after 60 reinforcements were delivered or 2 hours had passed. As training progresses the sequences of presses become more organized (Figure 53A and B), in parallel the distribution of sequence length shifts to the right (Figure 53C); the fraction of short (1 and 2 presses) sequences decreases from 0.66 to 0.27 (early - sessions 1 and 2; late - sessions 12 and 13) while longer sequence fractions increase from 0.12 to 0.36 (4 and 5 presses; from early to late sessions) and the efficiency improves from 11.2% (session 1) to 37% (session 13) ($p<0.01$, Figure 53D). The average sequence length quickly and significantly increases (sessions 1,2 v 5,6 or 12,13 $p<0.01$, Figure 53E) and stabilizes after session 5,6 while within sequence press rate monotonically increases ($p<0.01$, Figure 53F). The inter-sequence interval (ISI) decreases (sessions 1,2 v 5,6 $p<0.01$, Figure 53G) and stabilizes after sessions 5,6 while sequence duration decreases monotonically but not statistically significant ($p>0.05$, Figure 53H). Interestingly, while the CV of the sequence length and the CV of within sequence press rate quickly decrease and stabilize after sessions 5,6 (sessions 1,2 v 5,6 or 12,13 $p<0.01$, Figure 53I,J), the CV for ISI and the CV for sequence duration show a sharp significant decrease (sessions 1,2 v 5,6, $p<0.01$) followed by a rebound after sessions 5,6 (sessions 5,6 v 12,13, $p>0.05$, Figure 53K,L). Sequence length is a task relevant dimension while action duration and ISI are not. The increased CV observed for the 2 later non task relevant dimensions in sessions 12,13 is not statistically different from the values observed during sessions 1,2 ($p>0.05$, Figure 53K,L). Interestingly these rebounds in variability are consistent with the original results reported by Santos et al. (2015) which show that variability in task relevant dimensions is reduced (speed in the original data) while variability in orthogonal non relevant dimensions increases (length in the original data).

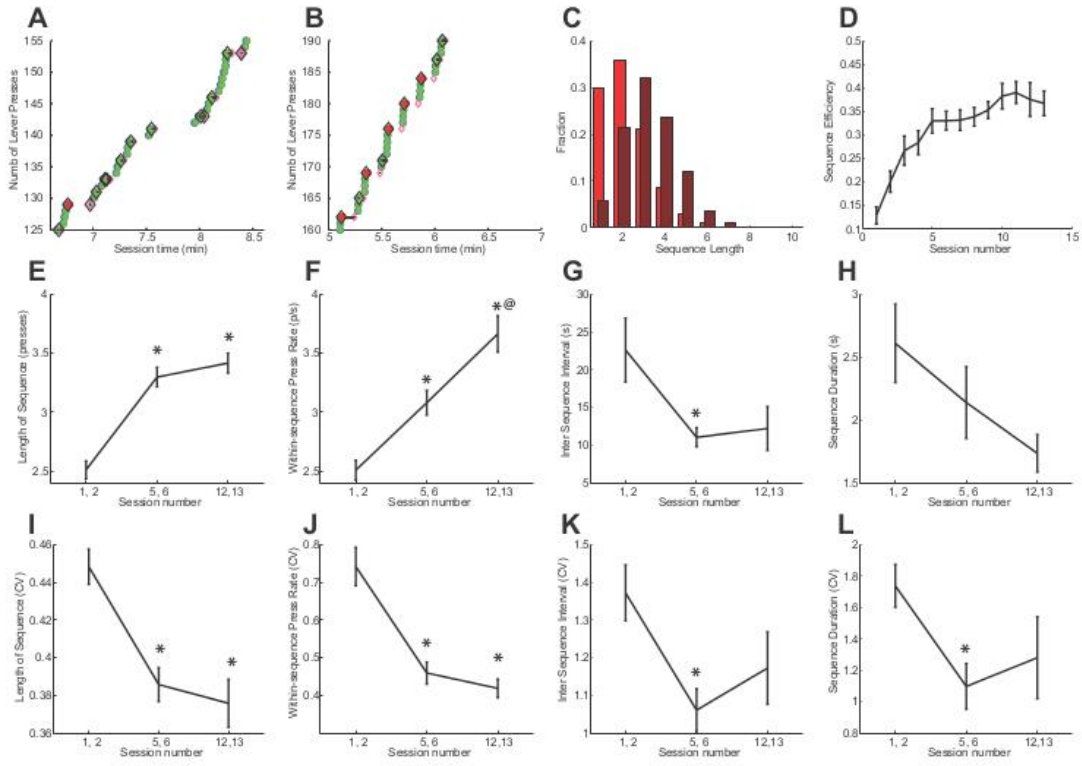


Figure 53: Detailed description of the accurate motor skill learning

A-B. Excerpts from session 1 (A) and session 13 (B). Blue circles show lever presses, green circles, leever releases and red circles show reinforcement delivery. Sequences become more organized as training progresses. Right shift in distribution of sequence length from early to late sessions (C). Increase in efficiency measured as (successful trials)/(all trials) (D). E-H, average number of lever presses per sequence (E), within sequence press rate (p/s) (F), ISI (G), sequence duration (H), during early (1,2), mid (5,6) and late sessions (12,13). I-L, Variability, measured as coefficient of variance (CV), of sequence length (I), within sequence press rate (p/s) (J), , ISI (K) and sequence duration (L), during early (1,2), mid (5,6) and late sessions (12,13). Error bars denote s.e.m. * denotes significance ($p<0.01$) compared to sessions 1,2 and @ denotes significance ($p<0.01$) compared to sessions 5,6.

Though mice can learn a complex motor skill which requires the end-point behaviour to be constrained within a strict range, it remains unclear if they keep track of their own performance. If so, could mice be trained to provide a report of how confident they are in their performance on a trial by trial basis?

Confidence estimation of action performance

In order to answer the previous question, an extra requirement was introduced in the task following the 13 training sessions reported above: the cached reinforcement was no longer presented immediately after mice reached the magazine, animals must wait inside the magazine for reinforcement delivery. If mice left the magazine and didn't come back before the waiting interval had passed, the trial was considered aborted (minor jitter was allowed during waiting time; mice could briefly leave and come back as long as the total time inside the magazine was superior to 50% of the whole waiting interval). Also, if mice were to press the lever while in those brief absences the trial was deemed aborted as well

(regardless of the duration of the absence). Mice were kept on a daily training regimen and the duration of the waiting interval was slowly increased until it reached 8 secs (Figure S2). All the other training parameters remained unchanged. In summary, target remains covert, reinforcement remains cached and the time waited in the magazine adds extra cost to the performance which drives animals to select the trials to which they should commit.

As training progressed the waiting interval is increased and, as expected, the overall fraction of aborted trials increased (Figure 54A). Interestingly, mice learned to abort trials contingent to their actions as a U shaped distribution centered around the covert target emerged (Figure 54B). Moreover, short incorrect sequences are aborted earlier in training than longer incorrect sequences likely due to the relative difference in occurrence. The absence of major changes in the distribution of sequence length (Figure 54C) during training while, in parallel, the fraction of aborted trials increases contingent to performance shows that the mice integrate the extra waiting time at the magazine port as an added cost to the original trial design. This drives trial abortion without prejudice to the execution of the lever press sequence.

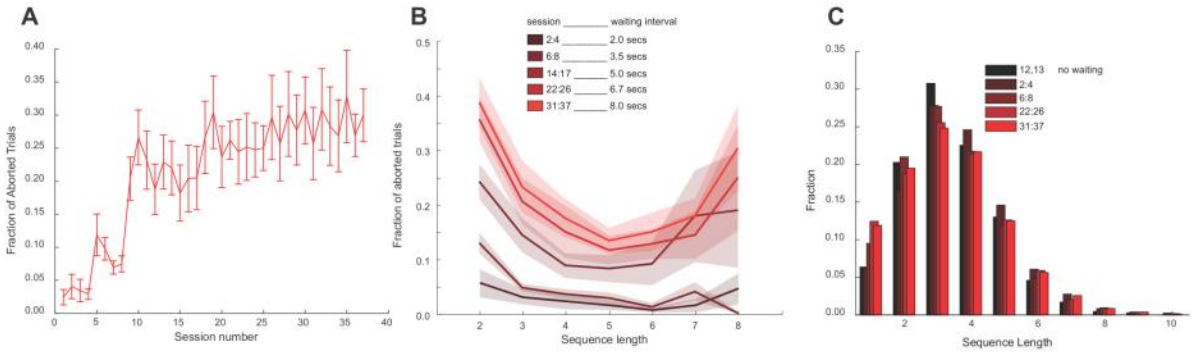


Figure 54: Emergence of the implicit confidence report

As the duration of the waiting interval increased, (A) mice aborted an overall larger fraction of trials. (B) The selection of aborted trials is contingent to performance. (C) No significant shift observed in distribution of sequence length. Color legends refer to the session number and progressively increased waiting interval.

After the waiting interval is extended to 8 secs the U curve becomes more symmetric and sharp (Figure 54B). The fractions of aborted trials for the target sequences (4 and 5 presses) are lower than the flanking incorrect sequences with either fewer or extra presses suggesting mice keep track of their current performance, compare it to a target and abort sequences contingent to the difference between the former and the later (Figure 55B). In order to better quantify the shape of the distribution of the fraction of aborted sequences a fit was performed using a quadratic polynomial as an approximation ($AX^2 + BX + C$; $R^2 = 0.98$); the parameter A corresponds to the curvature of the fit. Furthermore, a bootstrap analysis (100000 permutations) shows that the chances of such U curve occurring by chance are virtually zero (Figure 55C). Taken together the data and analysis show mice decided on abortion contingent to their performance. Therefore the fraction of aborted trials is considered as an implicit confidence report (Kepecs and Mainen, 2012b).

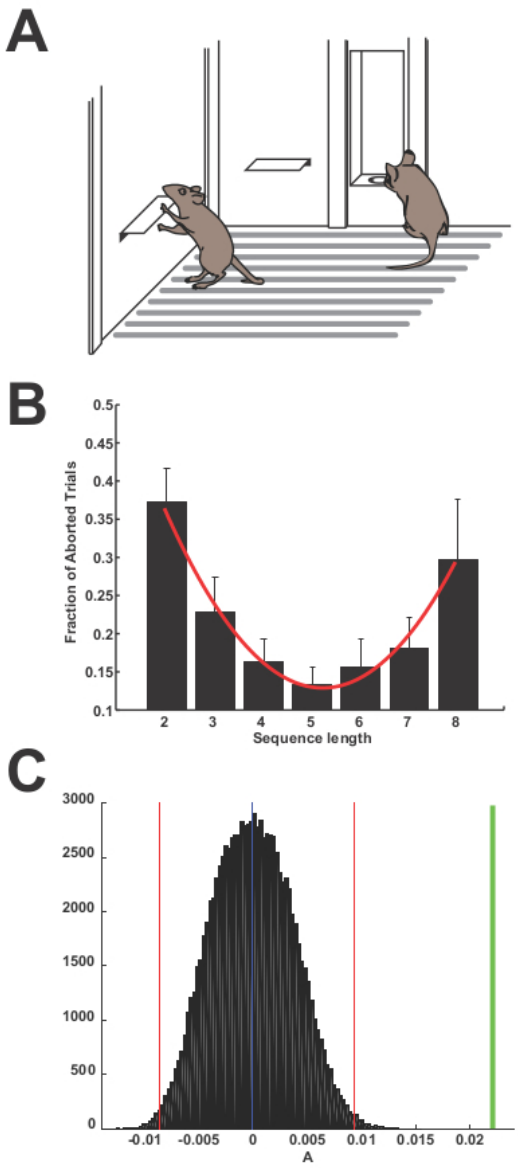


Figure 55: Accurate motor skill assay and confidence report

(A) Schematic representation of the operant box with lever and magazine where reinforcements are delivered. (B) Fraction of aborted trials as a function of sequence length. Bar graph shows the distribution of the mean of the fraction of aborted trials (with all sessions grouped together) by animal (error bars correspond to s.e.m. over animals). The red line is the quadratic fit to the data with an R2 = 0.98. (C) Comparison between experimental and permuted data shows the distribution of the curvature parameter of the quadratic fit derived from 100000 permutations of the original data. The permutation distribution is centered near zero (flat distribution of fraction of aborted trials vs sequence length, mean dark blue) away from the experimental value observed (light green). The odds of the experimental U shaped distribution happening by chance are statistically much smaller than 0.01 (red lines).

Finally, an important step in understanding how confidence is estimated is determining when it is calculated. It is unknown if mice compute confidence only when probed during report time (after they reach the magazine port) or before (in parallel with sequence performance or on their way to the magazine).

Confidence estimation timeline

Trials can be divided into 2 major epochs. The first encompasses the press sequence execution (from first lever press to last release) and includes the approach towards the magazine port. The second epoch starts when the animal performs the head entry which starts the clock and ends when the waiting interval is over or when the animal decides to leave the magazine (aborting the trial). If mice compute confidence before reaching the magazine, one could expect to see an analog (and early) confidence report in the latency to reach the magazine. In other words, one should expect faster movements (shorter latencies) when confidence is high and slower movements (longer latencies) when confidence is low.

The plot in Figure 56A shows (for all trials in the last sessions 31 to 37, waiting time = 8 secs) (black - aborted trials, red - committed trials) that while latency to check the magazine doesn't seem clearly different between aborted vs non aborted trials, the distribution of times waited in the magazine is bimodal: animals are more likely to either stay the whole waiting time or leave shortly after arriving (this is expected given that the wagering interval is constant; for a more detailed explanation see Kepecs and Mainen, (2012b). If instead just the means (+/- s.e.m.) are plotted for the same data (Figure 56B) it becomes clear that for aborted trials the animals move towards the magazine on average much slower than in committed trials ($p<0.01$). Interestingly, this is consistent with the strategy of minimizing the waiting cost by quickly aborting a perceived incorrect trial so that another can be started. Alas, examining the distribution of latency to check the magazine as a function of sequence length (Figure 57A), the U shaped curve observed in the previous confidence report (fraction for aborted trials vs sequence length, Figure 55B) is not mirrored in the latency to check ($R^2 = 0.86$). Furthermore, the odds of such curvature not occurring by chance are not statistically significant (Figure 57B).

It could be hypothesized that the magazine waiting time is very expensive for it requires a mandatory long delay added to immobility. Also it is an order of magnitude longer than the latency to check and therefore more relevant in controlling the flow of task execution (and maximizing the reinforcement rate), therefore clouding the latency to check (as an analog implicit confidence report) in the current task design. In order to better investigate this hypothesis another mice group (N = 12, BL6/C57) was trained in a similar task with one major difference: mice didn't have to remain inside the magazine; the waiting could happen anywhere inside the operant box; the reinforcement was delivered after the same delay (as in the original design). In order for a trial to be aborted mice had to, during the waiting time, perform a lever press (hence aborting the previous trial and initiating another). This new design keeps the delay for reinforcement delivery unchanged but relaxes the behavioral requirements for trial commit during waiting time. This variation in the task led to a much lower overall fraction of aborted trials (and no clear indication that fraction of aborted trials held information on performance monitoring) but yielded an analog confidence report in the latency to check the magazine. The curve observed in the distribution of latencies to check the magazine vs sequence length (Figure 57C) mirrors the U shape of fraction of aborted trials vs sequence length graph (Figure 55B). A fit analysis shows that the same polynomial approximation is perfectly adequate ($R^2 = 0.97$). In addition, a bootstrap permutation analysis (100000 permutations) shows the odds of such shape happening by chance are indeed statistically significant ($p<0.01$; Figure 57D).

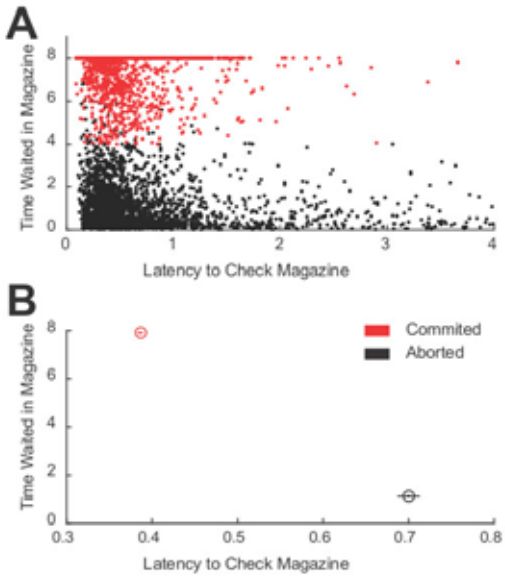


Figure 56: Latency to check as a confidence report

(A) Scatter plot of Time Waited in Magazine vs Latency to Check the Magazine (black - aborted trials; red - committed trials). (B) Same data as in (A) plotted as means (error bars are s.e.m. in both dimensions). Differences in latency to check the magazine between aborted and committed trials are more clear in B and statistically significant ($p<0.01$).

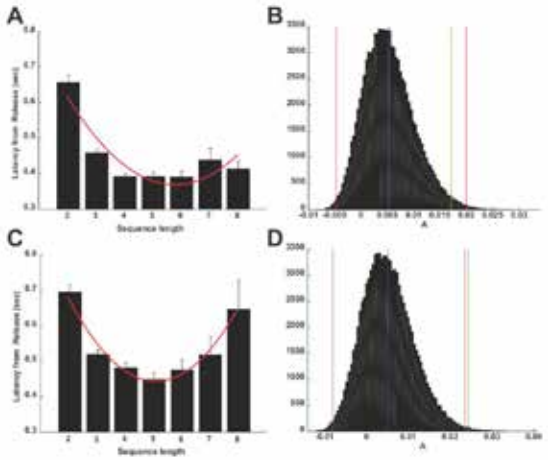


Figure 57: Two task designs: Differences in waiting time requirements for trial commit

(A) Latency to Check the Magazine as a function of Sequence Length for the original design. Bar graph shows the distribution of the mean of latency to check the magazine (with all trials grouped together; error bars correspond to s.e.m.). No clear indication of a U shape. (B) Permutation analysis shows that shape could happen by chance. (C) Same analysis for the task variation (see text for details). U shape Latency to Check the Magazine as a function of Sequence Length mirrors (Figure 1B) fraction of aborted trials vs sequence length. (D) The odds of the experimental U shaped distribution (green line) happening by chance are statistically much smaller than 0.01 (red lines).

Latency to check holds information that can be used as an analog report of confidence estimation before the time when trial abortion is probed. This is consistent with confidence being computed independent of (and ahead of) the report. While animals are capable of evaluating their performance on a fast trial to trial basis, it is not known whether variables other than the immediate performance could affect the decision to abort a trial.

Confidence estimation dynamics

When performing cognitive or motor tasks, animals regulate engagement depending on many variables: attention, satiation, exhaustion (Krebs et al., 2012; Salamone and Correa, 2012; Schouppe et al., 2014; Skvortsova et al., 2014; Varazzani et al., 2015; Verguts et al., 2015). If other variables than the immediate performance are pertinent for aborting a trial in the accurate motor skill task, further analysis might show distinctive temporal dynamics in the history of decision to abort. In order to dissect the dynamics of the decision to abort in a trial by trial basis, we performed a logistic regression (with lasso regularization, (James et al., 2013; Kass et al., 2014)) to probe the effect of performance and history of abortion in the current trial abort decision. The logistic regression analysis confirms that the performance in the current trial has a higher loading in the probability of aborting than previous (or future trials). It hints that another variable correlated to the history of abortion also predicts the probability of aborting (Figure 58A). The loadings on the history of abortion suggest a slower dynamics. A model where trial abortion is determined by 2 variables: 1) current performance (fast) and; 2) task engagement (slow), could account for the experimental data.

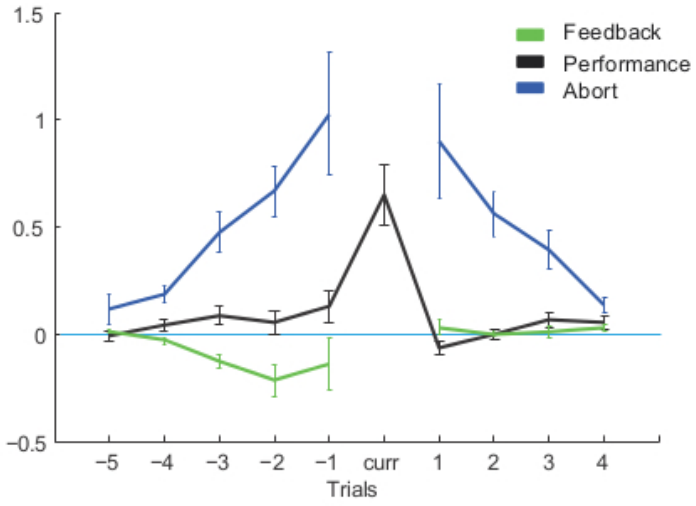


Figure 58: Logistic Regression analysis (traces of lagged trials)

(A) Logistic regression analysis of trial abortion probability. Logistic regression analysis predicting abort decision using different trial events (abortion, performance and feedback) at different lags (trials relative to predicted trial). The blue trace shows the average (over animals +/- s.e.m.) loading for abort decision (1 for aborted trials and 0 otherwise) on prior and future trials, i.e. the extent to which aborting on a given trial predicts aborting on nearby trials. The symmetric high loadings in previous and future trials suggests that a slow variable (i.e. engagement in task) is a strong determinant of the likelihood of current trial abortion. The black line shows the loadings for incorrect performance (of sequences; 1 for incorrect trials and 0 otherwise), i.e. the extent to which incorrect performance on a given trial predicts aborting on nearby trials. It has a large loading only on the current trial consistent with performance monitoring. Finally, the light green trace shows the loadings for feedback (which takes values of 0 on aborted trials, -1 in incorrect completed trials and +1 on correct completed trials). This variable aims to capture the effect of reinforcement obtained in correct vs incorrect sequences).

An alternative hypothesis is that a single variable affecting performance and abortion with the same dynamics could explain the data. If so, abort decision could be a direct consequence of low motivation (or lack of attention) and both performance and abort decision could be explained by a single underlying cause. We explore this hypothesis by numerically implementing a slow continuous variable (S) upstream to both performance and abort decision (Figure 59A). This is integrated in a system of logistic equations to

model, as binary variables, trial performance (P: correct vs incorrect) and trial abortion (A: committed vs aborted). Logistic regression can then be used to compare experimental and simulated time series of trial performance and abortion. The single variable model produces highly correlated time series for abort decision and performance (Figure 59B). This is translated in superimposed traces for A (blue) and P (black) in the logistic regression analysis (Figure 59C). This is due to: 1) the common slow component equally integrated in performance and abort decision and; 2) no performance integration in abort decision. If, on the other hand, the weight of the slow component on performance is decreased (formally equivalent to decreasing the correlation between S and P) and performance itself, is integrated in abort decision (equivalent to increasing the correlation between P and A), the simulation results change and the logistic regression analysis shows good agreement between simulation and experimental data (compare Figure 59D to Figure 58).

The simulation results show that a model where both decision to abort and performance are defined by a single upstream slow variable fails to approximate the experimental results (compare Figure 59C to Figure 58). On the other hand, a model where decision to abort (A) integrates 2 separate and uncorrelated variables, slow (S) and performance (P), correctly approximates the temporal dynamics of the time series for decision to abort (Figure 59D and Figure 58).

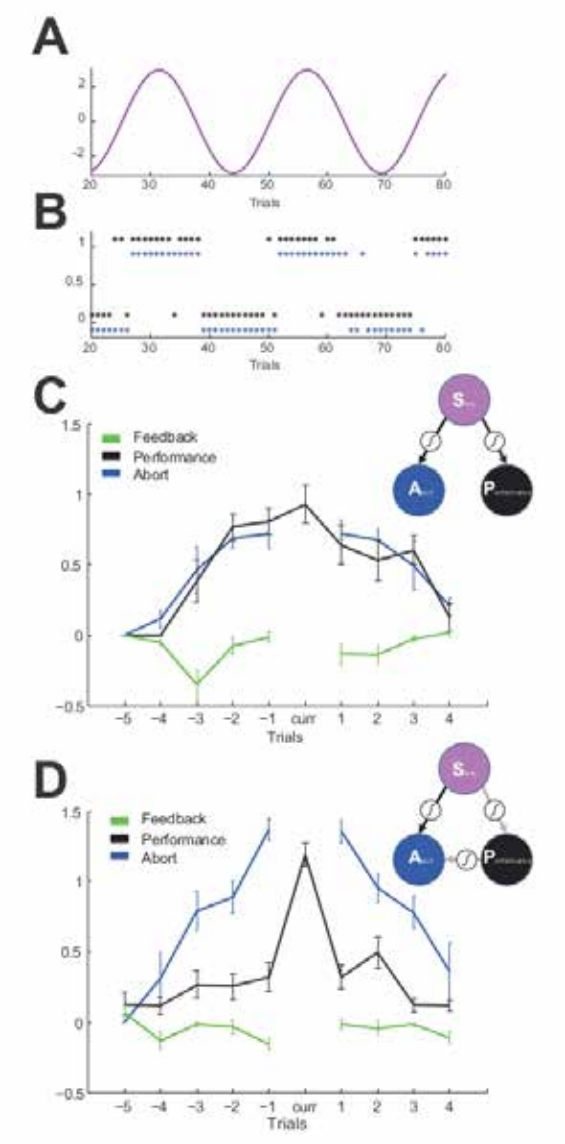


Figure 59: Single variable model and 2 variable equivalent model

(A) Slow variable time course of upstream variable. (B) Illustrative time series of the simulated abort decision (A, blue; 1 for aborted trials and 0 otherwise) and simulated performance (P, black 1 for incorrect trials and 0 otherwise) both determined by the upstream slow variable (S) in A (single variable model). (C) Logistic regression analysis for the single variable model data shows an overlap of the abort decision and performance incongruent with the experimental data. (D) Same analysis applied to a model with 2 uncorrelated variables (S and P) correctly approximate the experimental data (see Figure 2 for details).

Conclusions

In this report we describe a mice motor skill task with an implicit confidence report. Mice are capable of learning a self-paced operant accurate task with an end point constrained by strict upper and lower limits. The results also show, consistent with previous reports (Santos et al., 2015), that behavioural variability is differentially modulated depending on task relevance. Besides, when given the chance to wager on their most recent performance (by aborting trials), mice do so by selectively aborting trials contingent to error. This shows that mice keep track of their own performance and compute the difference to a target performance before feedback is provided (Figure 55).

The time spent in the magazine during the waiting interval follows a bimodal distribution: mice on average wait the whole interval (when committed) or leave shortly after arriving (in aborted trials). Furthermore, in addition to the decision to abort a trial when given the chance, mice approach the magazine on average faster when the decision to commit is selected and slower when abort is chosen. This suggests that the abort decision is defined before the mice reach the magazine and the report is collected. A similar analog (but explicit) report of confidence was observed in human's reaction time correlated with difficulty of a perceptual decision task (Kiani et al., 2014). One interpretation is that confidence estimation is more ubiquitous than originally expected and might be computed regardless of a trial structure where abort decision is probed.

When evaluating a just performed action, animals integrate variables beyond the temporal scope of execution. Such variables as motivation, attention, satiation and exhaustion might affect confidence estimation of actions. Logistic regression analysis shows that a in addition to a fast variable correlated to task performance, a slower variable correlated with trial abortion determines the choice to commit to a given trial (Figure 58). Computational modeling further reinforces the legitimacy of a 2 variable hypothesis. It remains unclear what are the neural substrates of the slow variable observed in the Logistic regression analysis.

Commonly referred as the sense of agency, the experience of being in voluntary control over one's own actions and, as a result, operating on the environment through goal-directed actions, is thought to be a defining human quality. The current report shows that animals are capable of monitoring their own performance. Moreover, the longer the waiting interval (analogous to the cost) the sharper the U shaped fraction of aborted trials vs sequence length. This result is consistent with previous data that support the hypothesis that effort increases the sense of agency (Demanet et al., 2013) but it is at odds with other data that suggest that the sense of agency is decreased when the contiguity between action and outcome is increased (which diminishes the temporal binding) (Moore et al., 2009).

This task paves the way for the investigation of the brain circuitry required for the calculation of confidence in action policies. The brain areas implicated in the estimation of confidence of action performance remain unknown. We are investigating the role of the Anterior Cingulate Cortex (ACC) previously implicated in error detection as well as the somatosensory system (especially S1 and M1) which have been implicated in sensory-motor integration. We predict that optogenetic inactivation of ACC would increase the overall fraction of aborted trials without changing the accuracy of the performance monitoring while inactivation of S1 will lead to a flat distribution of fraction of aborted trials, therefore a degradation of the ability of the animals for confidence estimation in action performance.

Data Provenance
The mouse data set on confidence estimation and action performance were collected at the Champalimaud Centre for the Unknown in Lisbon, Portugal by Rodrigo Freire Oliveira (post-doc at the Neurobiology of Action under the supervision of Dr. Rui M. Costa).

Data Location
Behavioral data were deposited on a server at
http://sp3.s3.data.kit.edu/3_2_1/Dataset2/SP3.mat

Completeness of data sets and models
The data collected so far includes all the behavioural experiments. The behavioural data were used to develop the first iteration of the model described in the text. Further development in the models is ongoing.

Data quality and value
The data collected shows the development of an implicit report of confidence estimation in action performance. It suggests that animals keep track of their own performance and decide on aborting ongoing trials based on a fast variable (depending on performance) and a slow variable (depending on motivation/engagement in task). The model simulation and analysis further reinforces this interpretation.

Publication: in preparation.

2.2 Mapping and understanding the neuronal circuits involved in motivation, emotion and reward

Task T3.2.2 - Mathias Pessiglione (ICM)

Overview

Mathias Pessiglione, Mauricio R. Delgado "The good, the bad and the brain: Neural correlates of appetitive and aversive values underlying decision making", Current Opinion in Behavioral Sciences. 5:78-84.

Abstract

Approaching rewards and avoiding punishments could be considered as core principles governing behavior. Experiments from behavioral economics have shown that choices involving gains and losses follow different policy rules, suggesting that appetitive and aversive processes might rely on different brain systems. Here we contrast this hypothesis with recent neuroscience studies exploring the human brain from brainstem nuclei to cortical areas. A strict anatomical divide seems difficult to draw, as appetitive and aversive stimuli appear to be processed in a flexible manner that depends on a context-wise subjective reference point. However, some valence specificity can be defined in the sense that net values (discounting appetitive by aversive values) are signaled with enhanced activity in some circuits, versus reduced activity in others. This dichotomy might explain why drugs or lesions can produce valence-specific effects, biasing decisions towards approaching a reward or avoiding a punishment.

Highlights

- The same brain regions process rewards or punishments across reinforcer modalities.
- No strict separation of brain systems processing appetitive and aversive events.
- Appetitive or aversive depends on a context-dependent subjective reference point.
- Some brain regions integrate appetitive and aversive aspects into net values.
- Net values are positively encoded in some brain regions, negatively in others.

Data set: Pharmacological manipulation of motivational processes

Dissociation of motor and motivational functions of dopamine in humans

Raphaël Le Bouc[1,2,3], Lionel Rigoux[1,2], Liane Schmidt[4,5], Bertrand Degos[2,6], Marie-Laure Welter[2,6], Marie Vidailhet[2,6], Jean Daunizeau[1,2], Mathias Pessiglione[1,2]

(1) Motivation, Brain & Behavior (MBB) Team, Brain & Spine Institute (Institut du cerveau et de la Moelle Epinière - ICM), Hôpital de la Pitié-Salpêtrière, 75013, Paris, France
(2) INSERM UMR 1127, CNRS UMR 7225, Université Pierre et Marie Curie (UPMC-Paris 6), Hôpital de la Pitié-Salpêtrière, 75013, Paris, France

(3) Urgences cérébro-vasculaires, Hôpital de la Pitié-Salpêtrière, APHP, Paris, France
(4) INSEAD Faculty & Research, Centre Multidisciplinaire des Sciences Comportementales Sorbonne Universités-INSEAD, 6 rue Victor Cousin, 75005 Paris

(5) Economic Decision-Making Group, Laboratoire des Neurosciences Cognitives, Ecole Normale Supérieure, Département d'Etudes Cognitives, 29 rue d'Ulm, 75005 Paris, France

(6) Département des Maladies du Système Nerveux, Centre Expert Inter-Régional de la Maladie de Parkinson, Hôpital de la Pitié-Salpêtrière, APHP, Paris, France

ABSTRACT

Motor dysfunction (e.g., bradykinesia) and motivational deficit (i.e., apathy) are hallmarks of Parkinson's disease (PD). Yet it remains unclear whether these two symptoms arise from a same dopaminergic dysfunction. Here, we developed a computational model that articulates motor control to economic decision theory, in order to dissect the behavior of 24 PD patients, tested On and Off dopaminergic medication, in motivation and choice tasks that both involved a trade-off between physical effort and financial reward. Model-based analyses in both tasks captured two differences, in reward sensitivity and movement dynamics, with two independent parameters, which predicted clinical improvement in apathy and motor dysfunction, respectively. We conclude that dopamine has independent roles in motivational and motor processes: it increases the amount of effort that subjects are willing to produce for a given reward, and accelerates the production of this effort irrespective of reward level.

SIGNIFICANCE STATEMENT

Many neurological conditions are characterized by motor and motivational deficits which both result in reduced behavior. It remains extremely difficult to disentangle whether these patients are simply unable or do not want to produce a behavior. Here, we propose a model-based analysis of the behavior produced in tasks that involve trading physical efforts for monetary rewards, so as to quantify parameters that capture motor dynamics and sensitivity to reward, effort and fatigue. Applied to Parkinson's disease, this computational analysis revealed two independent effects of dopamine enhancers, which predicted clinical improvement in motor and motivational deficits. Such computational profiling might provide a useful explanatory level, between neural dysfunction and clinical manifestations, for characterizing neuropsychiatric disorders and personalizing treatments.

INTRODUCTION

Why don't we make more effort? Is it because we don't want to, or just because we can't? This question is particularly hard to address in the case of patients with pathological conditions that combine motivational and motor deficits, such as Parkinson's disease (PD). Some of the motor symptoms that characterize PD, such as akinesia (paucituy of movement) or bradykinesia (movement slowness) are difficult to disentangle from apathy (motivational deficit), usually defined as a reduction of goal-directed behavior.

Candidate neurobiological mechanisms underlying motor and motivational deficits both involve dopamine. Motor symptoms (Rodriguez-Oroz et al., 2009) are primarily caused by the degeneration of dopaminergic neurons in the substantia nigra pars compacta (SNpc) that project on dorsal parts of the striatum(Ehringer and Hornykiewicz, 1960; Kish et al., 1988). Apathy, one of the most frequent non-motor symptoms in PD (Brown and Pluck, 2000; Marin, 1991; Starkstein et al., 1992), might also relate to dopamine depletion (Czernecki et al., 2008; Schmidt et al., 2008; Thobois et al., 2010), but more specifically to the degeneration of dopaminergic projections to the ventral striatum (Remy et al., 2005) arising from the ventral tegmental area (VTA) (Brown et al., 2012; Javoy-Agid and Agid, 1980). Thus, dissociation of motor and motivational deficits in PD requires a proper articulation of the putative roles of dopamine in motivation and motor functions, an

issue that has only recently received consideration in theoretical neuroscience (Rigoux and Guigon, 2012; Shadmehr et al., 2010).

Recent investigations of motor deficits have suggested that the kinematic characteristics of movements are preserved in PD, and that bradykinesia can be explained by a shift in a cost/benefit optimization process (Baraduc et al., 2013; Mazzoni et al., 2007). This optimization has been formalized in optimal control theory, which assumes that movement speed is adjusted so as to minimize a cost, related to the accuracy of the movement end point, or to the energy expended during movement execution (Guigon et al., 2007; Harris and Wolpert, 1998; Todorov and Jordan, 2002). However, the benefit is typically not taken into account in optimal control theory (Rigoux and Guigon, 2012), and consequently reward level has not been manipulated in previous investigations of motor deficits exhibited by Parkinsonian patients.

On the other hand, investigations of motivational deficits following dopamine depletion have generally neglected motor processes, even if testing motivation involves reading a motor output (Berridge, 2004). Incentive motivation can be construed as an implicit mechanism invigorating action execution in proportion to the expected reward, or as an explicit choice to exert more effort in order to get more reward. It has been formalized in economic decision theory, as well as in optimal foraging theory, as an optimization process that maximizes reward value while minimizing effort cost (Stephens and Krebs, 1986). The role of dopamine in promoting high effort – high reward behavioral policy has been well established in animals (Walton et al., 2006), and more recently evidenced in humans (Treadway et al., 2012; Wardle et al., 2011). However, the paradigms and analyses used in these seminal works did not allow specifying the pro-motivational effect of dopamine as either an enhancement of reward value or an alleviation of effort cost, which was also confounded with delay of reward delivery.

The aim of the present study is to provide a principled account of the motor and motivational functions of dopamine. For this, we developed a computational model that allows dissecting the effects of dopamine enhancers on the behavior produced by Parkinsonian patients in a cost/benefit trade-off task. The task was adapted from a behavioral paradigm in which payoff depends on a physical effort – the force exerted on a handgrip. More precisely, payoff was based on force peak, because it is a measure of action vigor that avoids confound with delay. This paradigm has been used previously in fMRI and lesion studies to demonstrate the implication of the ventral striato-pallidal complex in translating higher monetary incentives into greater physical effort (Pessiglione et al., 2007; Schmidt et al., 2008, 2012). Here, we implemented two versions of the paradigm (Figure 60), to capture both the processes that have been described as implicit invigoration and as explicit decision-making. The first version is an incentive force task, in which participants were free to exert any force between zero and their maximum, knowing that the payoff would be proportional to the force peak and to the monetary incentive, which was varied on a trial-by-trial basis. The second version is a binary choice task in which participants had to select either a variable high-reward/high-effort or a fixed low-reward/low-effort option.

Figure 60: Behavioral tasks

Successive screenshots displayed in one trial, with duration in milliseconds. **(A)** The incentive force task. After a fixation cross, subjects are shown the monetary incentive as a coin image (0.1, 0.2, 0.5, 1, 2, 5 €). This is the trigger to exert a force on a hand grip. Online feedback on the force produced is provided as a cursor moving up and down within a scale graduated from 0 to the maximal force of the subject. The height reached by the cursor determines the fraction of the monetary incentive earned in the current trial. **(B)** The binary choice task. After a fixation cross, subjects are shown two options side by side, each corresponding to a potential monetary reward (coin image) associated with a required force level (orange bar). Subjects select their preferred option on a keyboard and must then produce the associated force (i.e., raise the cursor up to the orange bar). In both tasks, cumulative total of monetary earnings is indicated at the end of the trial.

The model combines formalisms of optimal control theory, with minimization of the cost linked to the timing of force production, and decision-making principles, with optimization of the financial benefits relative to effort costs. When adjusted to the behavior observed in the two tasks through model fitting procedures, the free parameters give a computational profile for each patient and session. By comparing two sessions, one performed after 12-hour withdrawal of medication (Off state) and one performed one hour after last intake of dopamine enhancers (On state), we were able to specify the computational parameters that are under the influence of dopamine: namely reward sensitivity and motor dynamics. These two independent computational effects of dopamine enhancers respectively predicted the clinical manifestations of motivational and motor dysfunction.

RESULTS

Patients

Demographic data and clinical assessments have been summarized in Table S1. Patients (n=24) and controls (n=25) did not differ in terms of gender (13/12 vs 7/17, $\chi^2_{(47)}$=2.64, p=0.10), age (61.2 vs 57.0, $T_{(47)}$=1.63, p=0.11), or education (5.9 vs 5.2, $T_{(40)}$=1.15, p=0.26). As one could expect, we found higher apathy scores in Off-PD patients than in controls (Starkstein score: 14.3 vs. 5.2, $T_{(42)}$=6.63,

p<0.001). In PD patients, dopaminergic medication (either levodopa or dopamine receptor agonists) not only significantly decreased motor symptoms (UPDRS-III score: 32.8 vs. 11.7, $T_{(22)}$=9.75, p<0.001), but also significantly reduced apathy scores (Starkstein score 14.3 vs. 8.9, $T_{(16)}$=6.63, p<0.001) yet without normalizing them to controls (Starkstein score: 8.9 vs. 5.2, $T_{(44)}$=3.52, p<0.001). In the following, we report comparisons between controls and Off-PD patients to assess disease effect, and between Off-PD and On-PD patients to assess medication effect.

Table S1. Demographic and clinical data

	Controls (N=25)	PD on (N=24)	PD off
Age (years)	57.0 ± 2.1	60.2 ± 1.6	–
Sex (female/male)	13/12	7/17	–
Education level (seven points scale)	5.9 ± 0.3	5.2 ± 0.6	–
Disease duration (years)	–	11.4 ± 1.3	–
Dopaminergic daily dose (mg/day)	–	1148 ± 86	–
UPDRS-III motor score	–	11.7 ± 1.7	32.8 ± 3.2
Starkstein apathy score	5.2 ± 0.7	8.9 ± 0.8	14.3 ± 1.3
MADRS depression score	2.9 ± 0.7	5.9 ± 1.0	–
MMSE cognitive score	–	27.3 ± 1.3	–
Mattis dementia score	–	136.6 ± 14.3	–
Cued recall memory score	–	43.8 ± 5.0	–

UPDRS: unified Parkinson's disease rating scale; MADRS: Montgomery and Asberg depression rating scale; MMSE: mini mental state examination

Dopamine effect on force pulse dynamics

We first assessed how disease status and dopaminergic treatment affected motor contraction during force pulse, irrespective of monetary incentives (Figure 61A and B).

Controls and On-PD patients produced on average higher force peaks than Off-PD patients in both the force task ($T_{(43)}$=2.41, p=0.010; $T_{(19)}$=2.29, p=0.019) and the choice task ($T_{(43)}$=4.53, p<0.001; $T_{(19)}$=3.27, p=0.002). In addition, the force rise was faster, i.e. produced with higher peak of yank (the temporal derivative of force) in controls and On-PD compared to Off-PD patients in both the force task ($T_{(43)}$=4.51, p<0.001; $T_{(19)}$=2.51, p= 0.011) and the choice task ($T_{(43)}$=4.07, p<0.001; $T_{(19)}$=3.19, p=0.002). A similar difference was observed during the relaxation phase (Figure 61C): the decline in force also exhibited greater negative yank peak in controls and in On-PD compared to Off-PD patients ($T_{(43)}$=7.64, p<0.001; $T_{(19)}$=4.84, p<0.001). These slowing effects of dopamine depletion on contraction and relaxation yanks were correlated across patients (Pearson's rho=0.56, p=0.01; see Figure 61F). Below, we intend to demonstrate that the effects on force and yank peaks stem from independent motivational and motor functions of dopamine.

Dopamine effect on binary choice

We then tested whether PD and dopaminergic medication affect the amount of effort allocated to the different reward magnitudes in the choice task (Figure 60B), by looking at the indifference points obtained after convergence of the stair-case procedure. The three groups displayed a significant effect of incentives on choices (Figure 61D), meaning that they were willing to produce higher force peaks for higher incentives (all p<0.001). However, this effect (regression slope) was significantly reduced in Off-PD patients compared to both On-PD patients ($T_{(19)}$=2.19, p=0.021) and controls ($T_{(43)}$=1.76, p=0.039). Post-hoc comparisons showed that controls and On-PD patients chose higher force peaks than Off-PD patients specifically for the highest ($T_{(43)}$=3.51, p<0.001; $T_{(19)}$=2.76, p=0.006), but not for the lowest incentive level ($T_{(43)}$=0.41, p=0.341; $T_{(19)}$=0.50, p=0.312).

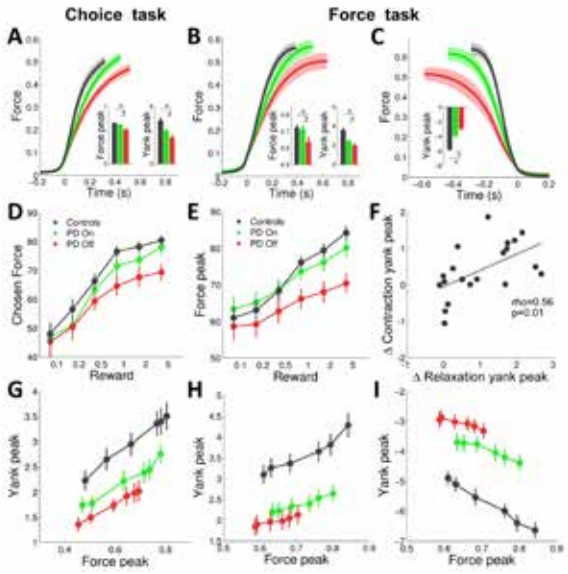


Figure 61: Behavioral results

(A,B,C) Average force dynamics for controls subjects (black), On-PD (green) and Off-PD (red) patients. Histograms show the average force peaks and yank peaks. Yank is the derivative of force with respect to time (df/dt). Force is expressed as a fraction of the subject-wise highest measure, and yank as this fraction per second. **(A)** Contraction phase in the choice task. **(B)** Contraction phase in the force task. **(C)** Relaxation phase in the force task. **(D, E)** Mean effects of incentives on selected forces (indifferent points) in the choice task **(D)**, and on produced force peaks in the force task **(E)**. Incentives are expressed in euros. **(F)** Correlation between dopaminergic effects on contraction and relaxation yank peaks in the force task. Each dot is a patient. **(G, H, I)** Scaling law relating yank peak to force peak. **(G)** Contraction phase in the choice task. **(H)** Contraction phase in the force task. **(I)** Relaxation phase in the force task. In all graphs, error bars are ± inter-subject SEM.

Dopamine effect on incentive motivation

Next, we tested in the force task (Figure 60A) whether disease status and dopaminergic medication affect incentive motivation, the process by which higher expected rewards are translated into greater efforts. The force peak significantly increased with incentive level in the three groups (Figure 61E, all p<0.001), but this effect (measured by the regression slope) was smaller in Off-PD compared to On-PD patients ($T_{(19)}$=2.11, p=0.024), and to controls ($T_{(43)}$=2.35, p=0.012). As in the choice task, post-hoc comparisons showed that control subjects and On-PD patients produced more force than Off-PD patients for the highest ($T_{(43)}$=4.15, p<0.001; $T_{(19)}$=2.92, p=0.004) but not for

the lowest incentive level ($T_{(43)}$=0.49, p=0.315; $T_{(19)}$=1.36, p=0.096).

Dopamine effect on motor scaling law

The preceding results indicate that dopamine has a similar role in the two tasks: it amplifies the weight of monetary incentives on effort production. Besides these motivational effects, motor effects were observed when examining the coupling of force kinematics parameters. Higher force peaks were linearly associated with greater yank peaks during both the contraction phase (Figure 61G and H, all p<0.001), and the relaxation phase (Figure 61I, all p<0.001). As shown in preceding analyses, the range of force peaks displayed by PD patients, in response to incentive levels, was narrower than in controls, and even more when Off compared to On. Yet the linear relationship between force and yank peaks was conserved in patients. Dopamine depletion manifested as a downward shift, meaning that equivalent force peaks were associated with lower yank peaks. This shift in the force-yank scaling law was significant in both the choice task (contraction phase, HC vs Off-PD: $T_{(43)}$=3.11, p=0.002; On-PD vs Off-PD: $T_{(19)}$=2.34, p=0.015) and the force task (contraction phase, HC vs Off-PD: $T_{(43)}$=4.79, p<0.001; On-PD vs Off-PD: $T_{(19)}$=1.92, p=0.035; relaxation phase, HC vs Off-PD: $T_{(43)}$=7.50, p<0.001; On-PD vs Off-PD: $T_{(19)}$=3.44, p=0.001). Thus, on top of the motivational effect of dopamine depletion that narrowed down the range of force peaks observed for the different incentive levels, a motor effect diminished the speed with which these force peaks were attained.

Computational analysis

We then studied how these dopamine-dependent modulations of effort production could be explained at the computational level. We developed a normative model that predicts how force dynamics should be selected in principle, depending on two contextual factors (incentive level and trial number) and four free parameters (reward sensitivity, cost sensitivity, fatigability and motor time constant: K_r, K_c, K_f, and τ). The predictions arise from a two-step optimization. The first step uses motor control equations to calculate the cost associated to each force peak (Figure 62, top). This estimation determines the dynamics of force rise over time (and therefore the yank peak): the one that minimizes the motor cost. The second step uses decision theory to calculate the net value (benefits minus costs) of each force peak (Figure 62, bottom). This valuation process determines which force peak will be produced in the force task, or selected in the choice task: the one that maximizes the net value. Note that one additional parameter was included in the model to fit the choices: this is choice temperature (beta), which captures the stochasticity of decisions.

We used simulations to verify that each parameter controlled a specific behavioral pattern. We then examined which free parameters best explained the effects of dopaminergic medication on choices and force and yank peaks. These effects could *a priori* be accounted for by a modulation of any of the four parameters. We considered the 2^4 possible combinations (modulation or no modulation for any of the four parameters). These 16 models were estimated and compared by families for each parameter using Bayesian model selection. The winning model was the one where dopaminergic medication affects both Kr and τ (with family exceedance probabilities xp>0.95), increasing the weight of monetary incentives and decreasing the time constant of motor contraction/relaxation (Figure 63A). This model provided a good fit for the three behavioral measures, *i.e.* force peak (mean R^2=0.94), choice (mean accuracy=0.70), and yank peak (mean R^2=0.92). We also tested whether the same parameters could account for the two tasks, by comparing this family of 16 models to an equivalent family with distinct sets of parameters for the two tasks. Although the latter better explained the data (xp>0.95), none of the parameters showed a consistent effect of task across subjects (all p>0.05). Moreover, we separately estimated the effects of dopamine in the two tasks, and found correlated estimates across patients for both Kr (rho=0.78, p<0.001) and τ (rho=0.43, p=0.049).

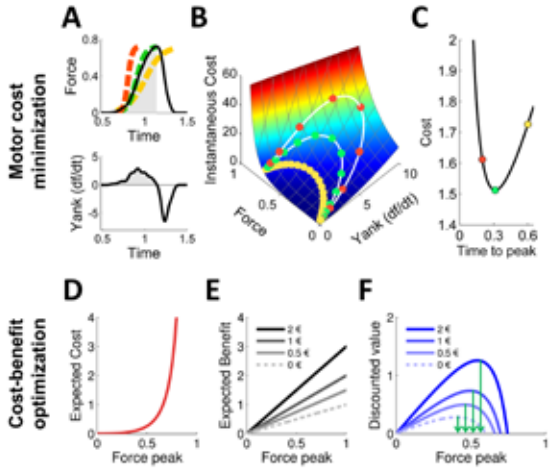


Figure 62: Computational principles

(A) Example of a force pulse (top), with force expressed in Newtons. The x-axis indicates time after trial onset in seconds. Yank (bottom) was calculated as the temporal derivative of force. Active periods of force pulses (in grey) were modeled as sigmoid functions (dashed lines) that approximate the solutions of an optimal motor-control model (see methods). The three colors correspond to three a priori possible trajectories in time. **(B)** Instantaneous cost (arbitrary units) for each value of force and yank. Simulated force-yank trajectories in time for three force pulses of decreasing duration (0.6, 0.3, 0.2 sec) and all reaching 70% of the maximal force are shown in the cost space (white lines). Circles indicate cost estimated at every 20ms step. The total cost of each force pulse is the integral of instantaneous costs across the duration of the active period of force pulse. **(C)** Total cost of force pulses (arbitrary units) as a function of effort duration (in seconds). This function defines the optimal duration (in green) that minimizes the cost of the force pulse. **(D)** Expected cost (arbitrary units) associated to every force peak, simulated at the optimal duration of force pulse (peak latency). Force peaks are expressed as a proportion of the maximal force. **(E)** Expected benefit (arbitrary units) is proportional to the incentive at stake and to the force peak in the incentive force task. Three possible monetary incentives are represented here (0.5, 1, 2€). The model assumes that force production is valuable in itself, even if the expected

financial reward is null. Because of this, the expected benefit can be superior to incentive level. **(F)** Net value associated to every force peak, for three incentives (0.5, 1, 2€).

Finally, we examined whether the model parameters could predict the clinical changes observed in patients after dopaminergic treatment. These clinical changes were improvement in both apathy and motor dysfunction assessed with the Starkstein apathy score and the UPDRS-III motor score, two effects that appeared unrelated across patients (*rho*=-0.11, *p*=0.66). Similarly, the effects of dopaminergic medication on *Kr* and τ seemed quite independent (*rho*=-0.12, *p*=0.69). Thus we tested the correlation between the changes in *Kr* and τ and the changes in apathy and motor dysfunction. We found that *Kr* modulation (*rho*=-0.52, *p*=0.02), but not that of τ (*rho*=-0.15, *p*=0.29), significantly predicted the alleviation of apathy, whereas τ modulation (*rho*=-0.44, *p*=0.03), but not that of *Kr* (*rho*=0.02, *p*=0.54), significantly predicted the improvement of motor dysfunction (Figure 63B).

DISCUSSION

In this study, we assessed the effects of dopamine depletion (comparing Off PD patients to healthy controls) and dopamine repletion (comparing Off to On PD patients) on effort allocation, using both binary choice and incentive force tasks. Model-free analyses showed that dopamine is causally involved in 1) amplifying the boosting effect of potential rewards on force production and 2) speeding up force rise to the peak, irrespective of expected rewards. We then developed a computational model of effort production in order to further characterize the dissociation of motivational and motor effects, focusing on the effect of dopaminergic medication in PD patients. Model-based analyses showed that dopamine enhancers increase reward sensitivity and decrease the time constant in motor drive, while leaving unaffected other parameters such as cost sensitivity, fatigability or choice temperature. In the following, we discuss these computational effects and their possible neural implementation.

Our results are consistent with the idea that dopamine helps with producing greater effort in order to obtain greater reward, an idea that has received a good wealth of evidence in animals (Salamone et al., 2012; Walton et al., 2009). Recent studies in humans have shown that *d*-amphetamine, a dopamine enhancer, enhances the willingness to exert effort (Treadway et al., 2012; Wardle et al., 2011). Here, we provide the first demonstration in humans that dopamine similarly enhances the propensity to select high reward / high effort options (in the choice task), and the energy actually invested in instrumental behavior (in the force task). These two processes could be considered as two different components of the behavior: orientation (which goal is selected) and intensity (how much energy is expended in goal pursuit). Our model nonetheless treats them as two instances of a same decision problem that consists of choosing a pair of effort and reward levels. The difference is that only two options are available in one task, whereas the option set is continuous (between 0 and maximal force) in the other task, making binary choice a special case of the incentive motivation problem. Yet from a psychological perspective, a crucial difference might be that the selected reward-effort pair is explicitly expressed before effort production in the binary choice but not in the incentive force task. This could change the behavioral output (force peak), as whenever possible the decision might be dynamically refined on the basis of sensory feedback. Our model is essentially static: it determines the best option on the basis of anticipated estimation of costs and benefits. Although it provided a good fit of force data in both tasks, the absence of dynamic adjustment might be one of its limitations.

Although previous studies did show that dopamine enhances the willingness to exert higher effort for higher reward, they did not disentangle between the possibilities that dopamine could increase reward attractiveness or decrease effort painfulness. Our model-based analyses suggested that the motivational effect of dopamine can be accounted for by an increase in K_r, the subjective weight of expected reward in the cost-benefit computation. This specific effect on K_r could also account for why dopamine helps overcoming various types of costs when seeking rewards, from effort (Salamone et al., 2012), to risk (St Onge and Floresco, 2009), or delay (Denk et al., 2005). The K_r effect is also consistent with the demonstration that midbrain dopaminergic neurons respond to stimuli that predict future rewards (Schultz et al., 1997), encode reward magnitude (Roesch et al., 2007; Tobler et al., 2005), and promote responses to reward-predicting cues (Arsenault et al., 2014; Tsai et al., 2009). Our finding also echoes a model suggesting that action vigor (i.e., frequency of lever press) is determined by the average reward rate per time unit, which would be encoded by dopamine level (Niv et al., 2006). This variable can be compared to our K_r parameter, which weighted the expected reward per force unit, and which was amplified by dopaminergic medication.

Conversely, dopamine did not change the fatigability parameter K_f nor the subjective weight of effort cost, K_c. These results are consistent with the absence of support for a role of dopamine depletion in fatigue (Willner et al., 1992), and with the observation that measures of nucleus accumbens dopamine or dopaminergic neuron activity are much more sensitive to expected reward than to expected effort (Gan et al., 2010; Pasquereau and Turner, 2013). Note that although costs were subtracted to benefits in our model, discounting was not linear. This is because the cost function was not linear but concave, following on the demonstration that perceived effort increases as a power function with force (Stevens, 1957). It has been recently shown that concave (parabolic) cost function provide a better fit of effort discounting than hyperbolic or linear functions (Hartmann et al., 2013). Divisive functions such as hyperbolic discounting are well adapted to delay discounting, since net values are kept positive, in accordance with the idea that an extremely delayed reward is still better than nothing. Yet this may not be true of effort discounting: climbing a mountain for a peanut may be worse than doing nothing. This is the reason why we opted for subtractive discounting, which allows for negative net values (worse than nothing).

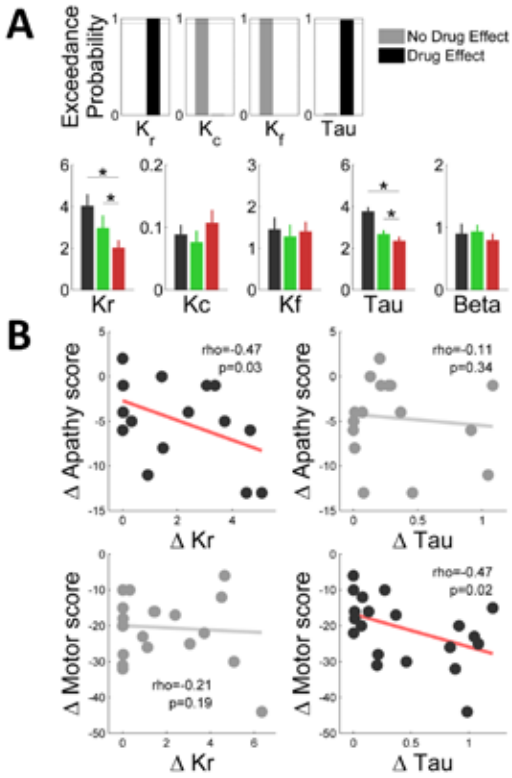


Figure 63: Computational dissection of dopaminergic functions

(A, top) Results of a Bayesian model selection comparing the plausibility of different possible modulations of behavioral response by dopaminergic medication in PD patients. For each parameter, the families of models with and without a modulation by dopaminergic medication were compared. Exceedance probability indicate how likely it is that one family is more frequent than the other in the population of PD patients. **(A, bottom)** Estimates of model parameters calculated at the session level in controls (black), On-PD (green), and Off-PD patients (red). **(B)** Correlation of medication effects on computational parameters *Kr* and *Tau* with clinical effects on apathy (Starkstein score) and motor dysfunction (UPDRS III score). Each dot is a patient.

Crucially, we found another effect of dopaminergic medication that was independent from reward level: after dopamine depletion, equivalent force peaks were produced with lower yank peaks, and this was independent from the restriction of force peaks produced for the different incentive levels. A similar shift in the motor scaling law had already been observed in PD patients performing a non-isometric task (Baraduc et al., 2013; Hartmann et al., 2013). Our results are consistent with the view that PD does not fundamentally change movement organization but restricts movement kinematics (Baraduc et al., 2013; Mazzoni et al., 2007). In our model, dopamine depletion decreased the time constant τ, which adjusts how motor drive impacts movement kinematics. Lower τ translates into slower muscle contraction for a given motor drive, and also slower relaxation. The mechanisms by which dopamine depletion slows down motor dynamics might involve an impaired selectivity in basal ganglia processing, leading to a failure to activate appropriate agonist muscles, or to inhibit antagonist muscles (Mink, 1996; Pessiglione et al., 2005). The modulation of τ and its consequences on contraction and relaxation slowness could therefore account for both bradykinesia and rigidity.

The motivational and motor functions of dopamine might be supported by topographically distinct functional networks, namely the mesolimbic and the nigrostriatal pathways. Apathy and motor symptoms might therefore reflect the heterogeneity in space and time of degeneration in PD. The dopaminergic loss occurs sooner and is stronger in the SN than in the VTA (Damier et al., 1999; Hirsch et al., 1988). This translates into a gradient with stronger dopamine depletion in the dorso-lateral putamen, compared to caudate and ventral striatum (Kish et al., 1988). On the one hand, the severity and asymmetry of motor symptoms in PD correlate with SN neurodegeneration (Du et al., 2012; Gorell et al., 1995), and with dopamine depletion in the dorsolateral striatum (Leenders et al., 1986), supporting involvement of the nigrostriatal pathway in motor dysfunction. On the other hand, the VTA has been hypothesized to play a key role in motivated behaviors (Tsai et al., 2009), through the mesolimibic projections to the NAcc (Berridge, 2007; Salamone et al., 2012) which has been conceived as a functional interface for translating motivational drives into motor or cognitive behaviors (Mogenson et al., 1980; Schmidt et al., 2012). Consistently, apathy in PD has been proposed to depend on the mesolimbic rather than on the nigrostriatal pathway, and consequently to dopaminergic denervation in the ventral striatum (Brown et al., 2012; Javoy-Agid and Agid, 1980; Remy et al., 2005; Thobois et al., 2010).

In conclusion, our computational analysis suggests that dopamine depletion down-weights expected reward in the cost-benefit computation, and thus lowers the acceptable effort costs, resulting in a reduction of goal-directed behaviors, *i.e* apathy. On top of this motivational deficit, dopamine depletion might also impair how acceptable costs are translated into movement kinematics, resulting in slower actions, *i.e.* bradykinesia. The motivational and motor effects of dopamine were captured by two distinct parameters of the model, which were correlated across patients to clinical assessments of motivational and motor deficits. We argue that computational phenotyping, i.e. the characterization of patients by model parameters adjusted on their behavior, might provide a useful intermediate explanation level between the clinical manifestations and the underlying neurophysiology. This computational approach could be applied to various pathological situations in order to help with personalizing treatments. In the present case, the two computational effects of dopamine are likely underpinned by distinct neural circuits, the mesolimbic and nigrostriatal pathways. Yet demonstrating such a link between computational parameters and underlying neural circuits would require further investigation.

METHODS

Force Task

The force task was designed to evaluate how subjects adjust their effort to incentive level. Subjects are instructed to try and win as much money as possible during the task, and are encouraged to perform as if they were playing for real money. The task includes 60 trials, corresponding to 10 repetitions of 6 monetary incentives (0.1, 0.2, 0.5, 1, 2, 5 €) presented in a random order. Each trial starts by the display of a fixation cross for 500ms. A monetary incentive then appears on the top left of the screen, presented as a coin or a bill image, simultaneously with a graduated scale (Figure 60A). The top line corresponds to producing the calibration force and winning the full incentive; each graduation corresponds to a fraction (10%) of the monetary incentive. Subjects are told that payoff was calculated as the fraction of the incentive proportional to the height they reached within the scale. They are provided with real-time visual feedback of the exerted force (with a cursor moving up and down within the scale). Appearance of the scale on screen is the trigger signal for subjects to start squeezing the handgrip so as to move the cursor up as high as possible, within a 5000ms interval. After every trial, the cumulative total of the money earned so far is displayed for 2000ms.

Choice Task

The choice task was designed to assess how subjects discount the value of reward prospects with the amount of effort that must be invested. Subjects are presented with a series of choices between low-reward/low-effort and high-reward/high-effort options. The low-reward/low-effort option is always presented on the left of the screen, and yields a reward of 0.05€ after exerting an effort corresponding to 10% of the subject's calibration force. The high-reward/high-effort option associates one of 6 possible rewards (0.1, 0.2, 0.5, 1, 2, 5 €) with a force level varying between 10% and 90%. Each option is presented as a coin or a bill image on top of a graduated scale with a red bar indicating the required force level (Figure 60B). Subjects decide whether or not it is worth exerting a higher effort to win a higher reward by pressing on the right or left arrow in a keyboard. The chosen option then remains on screen and the corresponding effort is implemented with the same visual display as in the force task. After every trial, the cumulative total of the money earned so far is displayed for 2000ms.

A staircase procedure was used to adjust the force level associated with every reward level in the high option, depending on subjects' choices, so as to gradually converge to indifference points, where subjects equally choose between the two options. At the beginning, the 6 possible rewards of the high option were respectively associated with efforts corresponding to 30/40/50/60/70/80% of the calibration force. After each choice, the effort level was either increased by 5% for the next occurrence of the same incentive, if the high option was chosen, or decreased by 5% in the opposite case. The task was made up of 15 repetitions of the 6 monetary rewards presented in a random order, for a total of 90 trials. This was sufficient to obtain a stable indifference point for each reward level.

Computational model

The basic principle of the model is a cost/benefit optimization (Eq. 1), where subjects intend to choose and produce the optimal force peak F, i.e. the one that maximizes a discounted value *V(F)*, calculated as the difference between expected benefits *B(F)* and expected effort cost $C(F)$. We opted for linear discounting to allow for negative net values, which accounts for the fact that doing nothing is sometimes better.

(1) $$V(F) = B(F) - C(F)$$

The benefit term *B(F)* was decomposed into reward-dependent and reward-independent components (Eq. 2). The reward-dependent component was by design proportional to the reward at stake *R* and to the exerted force *F* in the force task (see Fig. 3E), and only to the reward at stake in the choice task. It was weighted subject-wise by a free parameter K_r. The reward-independent component reflects the benefits of producing an effort, outside financial aspects, and was just proportional to the force level *F*.

(2) $$B(F) = \begin{cases} k_r RF + F & \text{(Force task)} \\ k_r R + F & \text{(Choice task)} \end{cases}$$

The expected cost *C(F)* was defined as the total motor cost *M(F),* multiplied by a subjective weight K_c and by a linear fatigue function (for the sake of simplicity), where *N* indicates the trial number and K_f the individual susceptibility to fatigue (Eq. 3).

(3) $$C(F) = k_c M(F)(1 + k_f N)$$

The total motor cost *M(F)* was defined after motor control theory. It was calculated as the integral of the instantaneous motor cost over the active period [0,*T*], i.e. from effort onset to force peak of an optimal force pulse (Eq. 4).

(4) $$M(F) = \min_u \int_0^T u(t)^2 dt, \quad [f(0) = 0, f(T) = F, \dot{f}(0) = \dot{f}(T) = 0]$$

Optimal force pulse (i.e the rising dynamics that minimizes the total motor cost for a given target force *F*) was modeled as a sigmoid function (see Figure 62A) that approximates the solution of an optimal motor-control model (Rigoux and Guigon, 2012) and requires lower computational resources. We defined the instantaneous motor cost as the quadratic neural drive $u(t)^2$, since motor control theory has shown that optimizing this cost minimizes the signal-dependent motor variability and reproduces the cardinal features of movement production (Guigon et al., 2007; Harris and Wolpert, 1998; Todorov and Jordan, 2002). The neural drive was calculated at each time point of the force pulse through a simplified model of muscular contraction (Eq. 5), in which the force dynamics $\dot{f}$ (the dot denotes the temporal derivative) was determined by the neural drive *u(t),* by a free parameter τ that individually adjusts the time constant of motor activation/deactivation, and by the current level of force compared to the maximal theoretical muscular force of the subject F_{max} (Eq. 5). F_{max} was modeled as another free parameter, superior or equal to the highest force produced by the subject. It was meant to reflect the total muscular mass, and was thus a priori unaffected by pharmacological manipulations.

(5) $$\dot{f}(t) = \tau u(t)[F_{max} - f(t)] - \tau f(t)$$

Equations (Eq. 3), (Eq. 4) and (Eq. 5) result in a cost function *C(F)* that links force peak *F* with its expected cost. This means that for each force level there is a unique motor cost, corresponding to the optimal (sigmoid) dynamics of force trajectory (the green one in Figure 62A, B and C). In other words, selecting a force peak automatically leads to selecting a trajectory in time, with a given yank peak, as reflected in the scaling law. Note that the cost function is explosive (Figure 62D): it goes to the infinite when subjects get closer to their maximal force. Thus, even if discounting is linear, the resulting net value function (Figure 62F) is not linear but follows an inverted U-shape, with a single maximum.

Thus, the free parameters of the model (*Kr, Kc, Kf* and τ) adjust the weights of objective quantities (reward, force and trial number) to individual susceptibility, in order to compute a subjective net value. The modeled net value was then used to predict the behavioral response on each trial of both choice and force tasks, with specific policy rules. For the force task, the predicted force peak was simply the argument that maximized the net value function (Eq. 6). In the choice task, there are only two possible reward and force levels. The decision was modeled with a softmax function (Eq. 7) that converted the difference in net value between the two options into choice probability, depending on a temperature parameter β. Finally, in both tasks, yank peaks were predicted by the equation of muscular dynamics (Eq. 4). Thus, the model was inverted by fitting three behavioral variables: choices, force peaks and yank peaks.

(6) $$F^{\star} = \underset{\mathrm{F}}{\operatorname{argmax}}\, V$$

(7) $$P_B = \frac{1}{1+e^{\frac{(V_A - V_B)}{\beta}}}$$

ACKNOWLEDGMENTS

The authors gratefully thank the patients for participating in the study; A. Welaratne, S. Aix, and the staff of the Neurology department for their help with management of patients.

2.3 Dissecting the brainstem modulation of cortical decision computations

Task T3.2.3 Tobias Donner (UVA), Andreas Engel (UKE)

Overview

Neuromodulation of cortical decision networks

Research over the past two decades has begun to uncover the mechanisms, by which people select actions based on a perceptual interpretation of their sensory environment. This process, termed sensory-motor or perceptual decision-making, has been studied extensively in non-human primates and more recently in humans. Typically, subjects are asked to judge weak sensory signals embedded in dynamic noise and to report their judgments with flexibly associated motor actions. At the algorithmic level, a key computation in such decisions is the gradual accumulation of multiple, noisy samples of "sensory evidence" for particular states of the world into a "decision variable" that forms the basis of action selection (Bogacz et al., 2006; Brunton et al., 2013; Gold and Shadlen, 2007; Ossmy et al., 2013; Ratcliff and McKoon, 2008; Usher and McClelland, 2001). The same computation is at play during higher-level decisions, such as choosing the option with the highest mean value from multiple streams of fluctuating numbers - a laboratory model of stock-market decisions (Busemeyer and Townsend, 1993; de Lange et al., 2010; Tsetsos et al., 2012; Yang and Shadlen, 2007). At the level of neurophysiological implementation, a large-scale network of cortical regions has been implicated in evidence accumulation (Figure 64). The posterior parietal and dorsolateral prefrontal association cortices (Donner and Siegel, 2011; Donner et al., 2007; Gold and Shadlen, 2007; Gould et al., 2012; Hebart et al., 2012; Heekeren et al., 2004; de Lange et al., 2010; O'Connell et al., 2012; Yang and Shadlen, 2007) seem to be the key nodes of this network. When choices are expressed as motor actions, pre-motor and motor cortices are also involved (Donner et al., 2009; Gold and Shadlen, 2000; Gould et al., 2012; de Lange et al., 2013; O'Connell et al., 2012; Wyart et al., 2012). Activity in all these cortical regions gradually ramps up during decision formation towards a critical threshold level, the crossing of which is invariably, and after a short delay, followed by the execution of the motor action. Biophysically detailed modelling has shown that the slow build-up activity is mediated by slow synaptic reverberation within recurrent cortical networks (Donner et al., 2007; Honey et al., 2012; Siegel et al., 2011, 2012; Wang, 2008).

Critically, each node of the large-scale cortical network depicted in Figure 64 is under the permanent influence of a number of modulatory neurotransmitters released from certain brainstem centers that send ascending projections to wide parts of the cortex (Figure 64, red). Very little is currently known about their functional impact on cognition, and, specifically, on evidence accumulation during decision formation. Three observations indicate that neuromodulatory brainstem systems might dynamically sculpt the cortical network interactions underlying decision formation, and, thereby, determine the internal state dependence of choice behaviour. First, it has recently become clear that some of these brainstem centers are *phasically* engaged (i.e., on a sub-second to second timescale) during rapid cognitive acts such as sensory-motor decisions (Aston-Jones and Cohen, 2005; Donner and Nieuwenhuis, 2013; Sarter et al., 2009) – causing rapid, cognition-linked elevations of physiological arousal. This stands in stark contrast to the traditional views of arousal systems as operating only on slow timescales and in an automatic fashion (Haider et al., 2013; Harris and Thiele, 2011; Steriade, 2000). Second, the modulatory neurotransmitters released from these brainstem centers control key circuit parameters in their cortical target networks (Aston-Jones and Cohen, 2005; Donner and Nieuwenhuis, 2013; Polack et al., 2013). Thus, neuromodulatory brainstem systems are in an ideal position to control the operating mode of the entire cortical network depicted in Figure 64

in a coordinated fashion. Third, dysfunctions of neuromodulatory systems go hand in hand with disturbed (and often less flexible) choice behaviour in some of the major neuropsychiatric disorders (e.g., schizophrenia or depression).

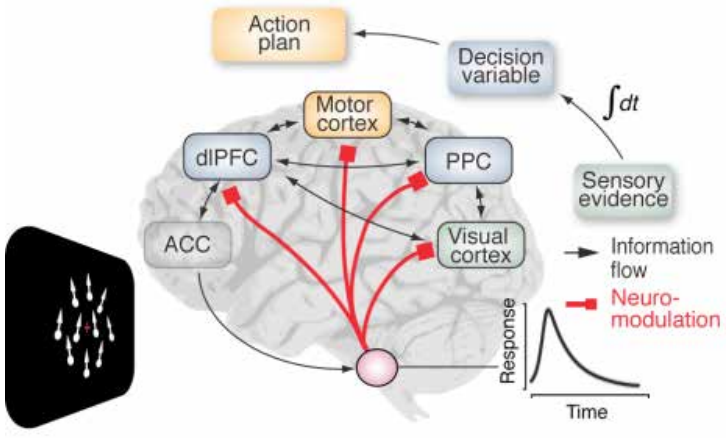


Figure 64: Neuromodulation of decision networks

During visuo-motor decisions (e.g. discrimination of weak visual motion signals), "sensory evidence" is accumulated and translated into an action plan. This process unfolds in a recurrent network of fronto-parietal and motor cortices. Each node is under the influence of ascending neuromodulatory brainstem systems, which activate phasically during decisions. The modulatory neurotransmitters released by these systems control the operating mode of cortical networks. PPC, posterior parietal cortex; dlPFC, dorsolateral prefrontal cortex; ACC, anterior cingulate cortex.

The goal of *Task T3.2.3 (Dissecting brainstem modulation of cortical decision computations)* is to uncover how cortical decision computations, the underlying cortical dynamics, and the resulting choice behaviour, are flexibly orchestrated by neuromodulatory systems – with a focus on the impact of these systems' phasic activations. In this *Overview Section 2.3.1*, we review recent work in two fields that form the backdrop of our project: Research into (i) the network dynamics of sensory-motor decisions in the human cortex and (ii) non-invasive monitoring of neuromodulatory transients in the human brain. In *Sections 2.3.2* and *2.3.3*, we then describe our recent progress on linking these two research lines by means of our work within the *HBP*.

Signatures of decision-making in the human cortex

Recent work on cortical decision dynamics using EEG and MEG in humans has employed sensory-motor decision-making tasks analogous to those extensively used in seminal single-unit recording studies in the macaque monkey cortex. This has helped identify two dynamical signatures of decision formation, which, in many ways, resemble the single-cell signatures of evidence accumulation in the macaque cortex (Gold and Shadlen, 2007): (i) choice-selective lateralization of beta-band power suppression in the cortical motor system (Donner et al., 2009; Gould et al., 2012; de Lange et al., 2013; O'Connell et al., 2012; Wyart et al., 2012) (Figure 65), henceforth referred to as "motor beta-lateralization"; and (ii) sustained magnetic fields or electrical potentials over parietal cortex (de Lange et al., 2010; O'Connell et al., 2012), henceforth referred to as "slow parietal potentials".

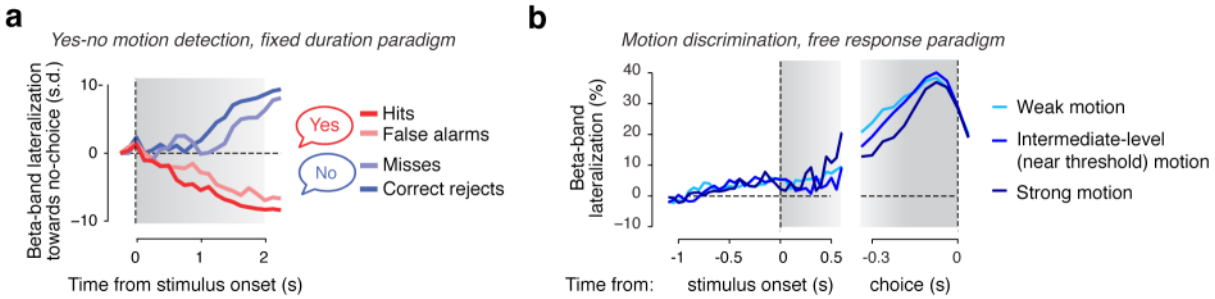


Figure 65: Build-up of movement-selective motor beta-lateralization during decision formation

a. Build-up of movement-selective activity in primary motor cortex during decisions about the presence or absence of coherent motion signals (2 s stimulus duration, followed by delay). Traces show the lateralization of beta-band (12-36 Hz) MEG power in motor cortex (contralateral – ipsilateral to hand indicating "yes"), which ramps up in opposite directions before "yes" and "no" choices and reflects the integral of the stimulus responses in motion-sensitive visual cortex. Adapted from ref.(Donner et al., 2009). **b.** Similar build-up during free response motion discrimination task. Subjects judged the direction of random dot motion. Traces are beta-band lateralization for different levels of motion coherence, pooled across choices (always ipsilateral – contralateral to chosen hand). Adapted from ref (de Lange et al., 2013).

Motor beta-band lateralization is a well-known dynamical signature of selective motor preparation. Recent work has established that this signature shares the hallmark functional properties with the choice-selective patterns of single-cell activity observed in several cortical areas of the macaque brain (Figure 65). When choices are indicated with left or right hand movements, the lateralized suppression of beta-band power (contralateral<ipsilateral to upcoming movement) ramps up as a cumulative function of the instantaneous sensory evidence encoded in visual cortex; when the decision is then maintained in working memory for an instructed delay before response execution, the beta-lateralization also maintains an elevated amplitude (Figure 65a) (Donner et al., 2009). When different mean levels of evidence strength (e.g. motion coherence) are presented on the screen, the beta-lateralization builds up at different rates (Figure 65b) (de Lange et al., 2013). When averaging the beta-lateralization time-locked to response in free-response reaction time tasks, the signals converge on a common level just before execution of the response, regardless of evidence strength (Figure 65b); this is consistent with an accumulation-to-bound mechanism (de Lange et al., 2013). Finally, and importantly, the motor beta-lateralization is selective for the movement choice (Donner et al., 2009; de Lange et al., 2013). Consequently, the experimenter can use it to predict, on a trial-by-trial basis, the specific upcoming choice of the subject (as quantified by a metric termed "choice probability" (Donner et al., 2009)) again just like single-unit activity in macaque cortex (Gold and Shadlen, 2007).

The slow parietal potentials, by contrast, are not selective for the specific choice: they ramp up invariably of the identify of the evidence or the evolving the choice (de Lange et al., 2010; O'Connell et al., 2012). Two other features make them an informative complement to the motor beta-lateralization: (i) they are generated in parietal association cortex, rather than motor cortex; (ii) they reflect the decision process even when that process is decoupled from action planning, and, therefore, not evident in the motor beta-lateralization (O'Connell et al., 2012). No signature with the latter property has yet been identified in the monkey brain. Because of their non-selective nature, the slow potentials might reflect a general confidence signal (i.e., reflecting the absolute value of the signed decision variable) and dynamically builds up, along with the unfolding decision itself. In sum, motor beta-lateralization and parietal potentials have complementary functional characteristics and should thus be tracked concurrently in future research.

Both electrophysiological decision signals above reflect local cortical dynamics that originate from specific nodes in the large-scale cortical network depicted in Figure 64. In that respect, they both differ from a third cortical signature of decision-making identified in other work: large-scale beta-band oscillations that are distributed (and coherent) across fronto-parietal association cortices (Figure 66). Coupled fronto-parietal beta-band oscillations have been found in various contexts during decision formation, in both humans and monkeys (Donner and Siegel, 2011; Donner et al., 2007; Gross et al., 2004; Haegens et al., 2011; Hipp et al., 2011; Pesaran et al., 2008; Siegel et al., 2012). The fronto-parietal beta-band oscillations differ in terms of several functional characteristics from the motor preparatory beta-band activity in motor cortex – for example, the strength of beta oscillations in fronto-parietal association cortex is enhanced, as opposed to suppressed, during decision formation. These oscillations may reflect network reverberations or neuromodulatory effects during the decision (Siegel et al., 2011).

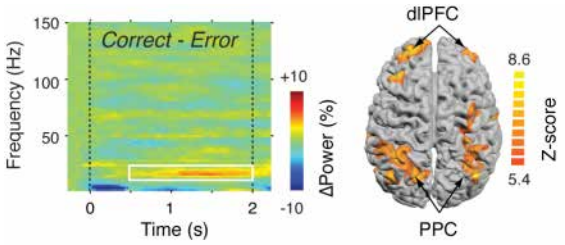


Figure 66: Fronto-parietal beta-oscillations

MEG power in the 12-24 Hz range during a motion detection task in dlPFC and PPC is larger before correct choices than before errors. Left, time-frequency representation of power difference. Vertical bars, stimulus on- and offset. The beta-enhancement (white box) occurs during decision formation. Right, source reconstruction of beta power difference. Adapted from ref. (Donner et al., 2007).

Decision-related neuromodulatory transients

Many neuromodulatory systems are thought to be involved in *some* aspect of decision-making (e.g., the dopamine system in learning action values) (Dayan, 2012). However, several lines of evidence implicate one system in particular as an important "orchestrator" of the cortical network dynamics underlying evidence accumulation in sensory-motor decisions: the noradrenergic locus coeruleus (LC-NA) system (Aston-Jones and Cohen, 2005). The LC receives descending, top-down projections from frontal regions that are connected to the cortical network (such as the ACC; Figure 64) and sends widespread projections to the entire cortical network from Figure 64, including visual and parietal cortex (Sara, 2009). Recent single-unit recordings from monkey LC revealed phasic responses that are specifically linked to sensory-motor decisions (Aston-Jones and Cohen, 2005). In other words, this low-level brainstem center is continuously informed about ongoing decision processes. NA released from the LC, in turn, controls the neural noise levels and the gain of synaptic interactions in the cortex (Aston-Jones and Cohen, 2005; Donner and Nieuwenhuis, 2013; Polack et al., 2013); on longer timescales, NA seems to be an important enabling factor for cortical plasticity mechanisms (Roelfsema et al., 2010). For these reasons, *Task T3.2.3* focuses on the LC-NA system. This focus does not exclude the possibility that other neuromodulators, (e.g., dopamine or acetylcholine) might also exhibit phasic activations during decisions and play similar roles in the decision computations studied here – in fact, we are also probing into some of these other systems. Monitoring transient modulations in the activity of these brainstem systems non-invasively in humans is important, not only for studying their basic functions in the healthy brain, but also their disturbances in several important neuropsychiatric disorders.

Such non-invasive monitoring can be performed in two complementary ways. First, the activity of specific brainstem centers can be tracked with fMRI (D'Ardenne et al., 2008;

Iglesias et al., 2013; Payzan-LeNestour et al., 2013). fMRI of the human LC is possible in principle, but it is technically challenging, due to the LC's small size and vicinity to the 4th ventricle (a source of artefacts). Consequently, initial attempts using standard fMRI procedures for LC-fMRI have attracted a lot of attention, but were met with skepticism (e.g., ref. (Astafiev et al., 2010)). A substantial part of *Task T3.2.3* has been devoted to developing an fMRI approach that overcomes these challenges (see *Section 2.3.1*).

Second, neuromodulatory activity might be monitored indirectly via changes in pupil diameter at constant illumination. Indeed, recent studies combining pupillometry with advanced neurophysiological techniques in rodents have shown that ongoing pupil diameter closely tracks ongoing fluctuations in global cortical state – which, in turn, are thought to be caused by ascending neuromodulatory systems (McGinley et al., 2015). In line with these observations, anatomical and physiological evidence points to a close anatomical and functional link between neural activity in the LC and the peripheral apparatus controlling pupil diameter (Aston-Jones and Cohen, 2005; Joshi et al., 2016; Loewenfeld, 1993; Nieuwenhuis et al., 2011). Similar links may exist for other neuromodulatory centers. Thus, fluctuations of pupil diameter can be used as a peripheral index of the brain's neuromodulatory state.

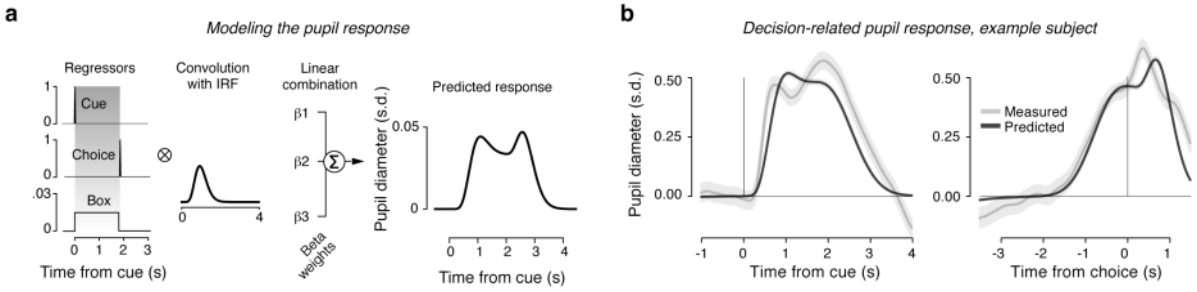


Figure 67: Modeling pupil dilation response during a visual decision task

a. Illustration of GLM. Left: Three temporal components that might drive the decision-related pupil dilation during a near-threshold contrast detection task in the presence of dynamic noise. Middle, pupil impulse response function (IRF) and best fitting beta weights. Right, predicted response for example trial. **b.** Mean decision-related pupil response and s.e.m. (grey), aligned to decision onset and to choice. Black lines, GLM prediction. Adapted from ref. (Gee et al., 2014).

Several on-going studies in our laboratory aim to characterize in detail how non-luminance-mediated pupil dynamics track perceptual decision processes (e.g., Figure 67). One key advantage of this pupillometric state index is that it can be seamlessly combined with electrophysiological recordings of cortical decision dynamics afforded by, e.g. MEG. In *Task T3.2.3*, we combined pupillometry with fMRI in *Data Set 1* to pinpoint the specific brainstem regions underlying pupil dilations and map out the associated changes in cortical state; we combined pupillometry with MEG in *Data Set 2* to identify the impact of neuromodulatory state on the decision-related cortical network dynamics described in this section above.

One influential account of the role of the phasic LC-NA release in sensory-motor decisions (Aston-Jones and Cohen, 2005) holds that the phasic LC-responses are triggered by the threshold-crossing of one accumulator in the cortex (the distributed nature of the accumulation process implies that not all "accumulator regions" might reach that threshold at the same time); the LC then broadcasts the decision commitment throughout the brain and the resulting cortical gain enhancement induced by NA then facilitates (speeds up) the resulting motor act. An alternative possibility is that the LC-NA release already occurs *during* the evidence accumulation process preceding threshold crossing. If so, LC-NA release would be in a position to alter the accumulation dynamics as the decision unfolds. Recent work from our lab provides initial indirect evidence for that

second alternative (Gee et al., 2014). We found that the neural drive of pupil dilation during perceptual decisions does not just occur at the final behavioural choice, but already throughout the preceding decision process (i.e., a robust contribution of a sustained “box” regressor across the decision interval in the simple linear model of the central pupil drive illustrated in Figure 67). This finding, in turn, suggests that phasic LC-NA release can shape the evidence accumulation process while the decision unfolds. One specific aim of *Task T3.2.3* is to arbitrate between the two above alternative scenarios: (i) *intra-decisional* LC-activity that shapes the accumulation process, and (ii) *post-decisional* LC-activity that controls only the decision threshold and/or motor latency.

Data set 1: Pupil-linked brainstem responses and the computation of yes vs. no decisions (fMRI)

We first aimed to model the functional impact of these decision-related, phasic elevations in pupil-linked arousal within the theoretical framework of evidence accumulation (also referred to as sequential sampling). Further, we aimed to pinpoint their neural correlates at the level of the brainstem and the cortical networks.

Behavioural task

To this end, in both *Data Set 1* (fMRI, described in this *Section*) and *Data Set 2* (MEG, described in *Section 2.3.3*), subjects performed the same yes-no visual choice task as used in our previous pupillometry work (Gee et al., 2014) that had revealed protracted pupil drive during visual decision-making (Figure 67). On each trial, subjects viewed a flickering stream of dynamic white noise and judged the presence ("yes") or absence ("no") of a target signal (a contrast grating) superimposed onto the noise (Figure 68). Before the main experiment, the signal contrast was titrated to each individual's 75% correct level. The signal grating, if present, was oriented either 45° clockwise or counter-clockwise; orientation was constant within each scanning run (comprising 40 trials). The dynamic noise was continuously present from the onset of the decision interval (cued by a tone) to keep overall luminance constant across all trials. The signal was present on half the trials. Once sufficiently certain (free response protocol; deadline: 3 s), subjects pressed a button with the left or right hand. The choice (button press) was followed by an inter-trial interval between 4 and 12 s (uniform distribution). This slow, event-related design was chosen to ensure that we could reliably characterize transient fMRI responses even in subcortical brainstem structures with unknown hemodynamics.

Data set

Fifteen healthy subjects (5 females; age range: 22–29 years) participated in the fMRI study. Each subject participated in three scanning sessions, on different days: one to define retinotopically organized cortical visual areas (about 75 min per session) and two sessions to measure fMRI responses in the main experiment (about 120 min per session). One subject was excluded from the analyses because the stimulus software did not receive the triggers from the MRI scanner in two out of three scanning sessions.

Analysis of pupil data

We computed task-related pupil response (TPR) on each trial as the mean of the pupil values in the window −1 s to 1.5 s from choice minus the mean baseline pupil value during the 0.5 s before trial onset. Trials were then sorted by TPR, after removing (via linear regression) the effect of signal presence and variations in reaction time (RT). The latter was important for the analysis of fMRI data, but we verified that all pupil-linked behavioral effects reported below are also evident without removing these components (data not shown). For each subject, we pooled trials into three bins containing the lowest and highest 40%, as well as the intermediate 20%, of TPR values (Figure 68d,e). This achieved a trade-off between maximizing both (i) trial counts in the high and low TPR bins and (ii) the disparity between the TPR amplitudes for both bins. In fact, this procedure yielded, on average, negative responses (i.e., task-related pupil constrictions) for the low TPR bin (Figure 68d,e).

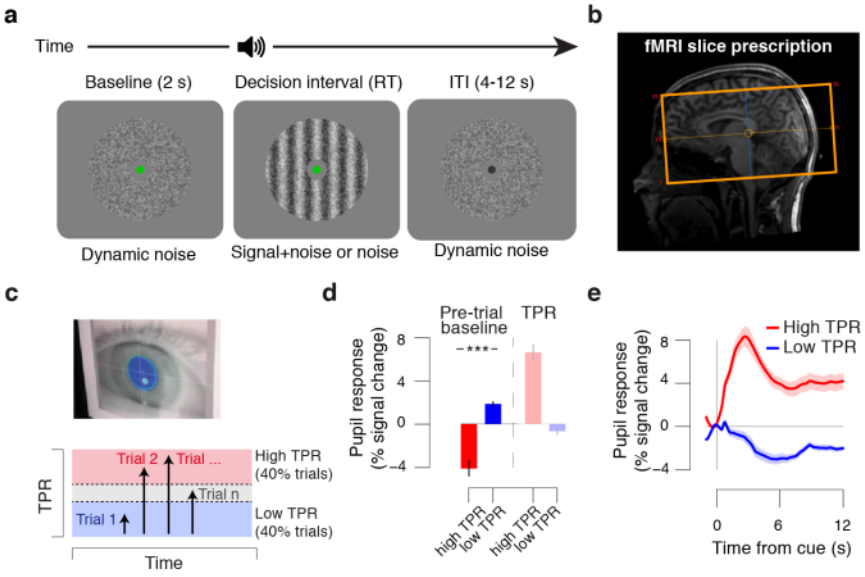


Figure 68: General approach

(a) Behavioral task. Example signal-present trial. The target grating, if present, was superimposed onto dynamic noise, until the subject's choice (button press). Signal contrast is exaggerated for illustration purposes; in the experiment it was low (75% correct threshold). (b) fMRI slice prescription. (c) Binning of trials by task-related pupil response (TPR) amplitude. (d) Average pre-trial baseline pupil size and TPR for high and low TPR trials. (e) Average pupil response time courses for high and low TPR trials.

Modeling the pupil-linked modulation of behaviour

To model the behavioural effects of the TPR-linked fluctuations in central arousal state, we fitted the drift diffusion model (Ratcliff and McKoon, 2008) to RT distributions for yes- and no-choices ("response coding"), separately for high and low TPR (Figure 69b). The model can be regarded as a dynamic version of signal detection theory. It posits the accumulation (without leak) of noisy sensory evidence into a decision variable. Once the decision variable reaches one of two bounds (here: for "yes" and "no"), the corresponding is made. The model decomposes the RT distributions and choices into a set of parameters that describe different mechanistic elements of the decision process at the algorithmic level: (i) The mean drift across trials is the "drift rate". (ii) The "drift criterion" refers to a bias in the accumulation process toward one or the other bound, irrespective of the evidence. (iii) The trial-to-trial variability in drift rate ("drift rate variability") is a separate parameter necessary for obtaining good fits to empirical RT distributions (Ratcliff and McKoon, 2008); this parameter is difficult to estimate and was assumed to be constant across TPR bins. (iv) The distance between both decision bounds - "boundary separation" - dictates how much evidence must be accumulated until a choice is made. (v) *"starting point"* determines a pre-potent tendency toward one or the other bound before the start of the decision process. (vi) The "non-decision time" lumps together the latencies of the sensory encoding and response execution processes preceding and succeeding the decision.

To obtain robust individual parameter estimates despite a comparably low number of trials (due to the slow event-related fMRI design, see above), we used the python toolbox HDDM (version 0.6)(Wiecki et al., 2013) to fit the model to a large group of subjects (total: *N* = 32) performing the yes-no contrast detection task including those from ref. (Gee et al., 2014) and the current fMRI study. Importantly, this was only used to obtain robust group-level priors on the parameters; the statistical comparisons between the individual parameters from high and low TPR bins in Figure 69c are done within the group of the current fMRI data set.

We found that subjects were conservative in their choices (i.e., inclined to choose "no", regardless of the evidence), and that this conservative bias was reduced on high TPR trials, while their sensitivity was unaffected. This effect was evident in either signal detection-theoretic measures (i.e., *criterion* and *d'*, Figure 69a), or the drift diffusion model parameters (Figure 69b,c). In particular, diffusion modeling revealed that subjects' accumulation process was generally biased toward "no", but, under high TPR, this bias was reduced, approaching 0 (Figure 69c), as required by our task to optimize performance. Importantly, we found no effects of TPR on any other model parameter – in particular, neither the drift rate (sensitivity of the decision process), nor the boundary separation or non-decision time. These results are in line with the idea that pupil-indexed arousal shapes the dynamics of the evolving decision, as opposed to affecting only post-decisional processes. In the asymmetric (i.e. yes vs. no) choice task studied here, pupil-linked arousal predicts a selective reduction of a conservative accumulation bias, which sets important constraints on the changes in cortical decision processing that might mediate this effect.

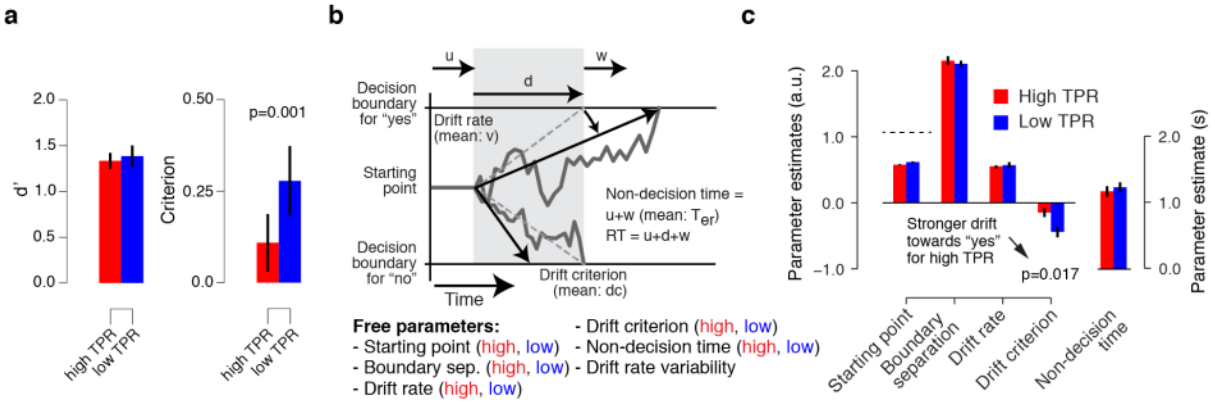


Figure 69: Pupil-linked modulation of behaviour

(a) Signal detection theory measures d' and criterion separately for high and low TPR trials. (b) schematic of the drift diffusion model as was fitted to the behavioural data. (c) Fitted DDM parameters separately for high and low TPR trials. The dashed line indicates a neutral starting point (half the boundary separation).

To gain deeper insights into the potential mechanisms that might mediate the above pupil-linked effects, on-going work simulates an abstracted neural ("leaky competing accumulator") model of the yes-no decision under fluctuating levels of neuromodulation. The model consists of a neural population encoding the yes-choice and another encoding the no-choice, whereby each population is modelled as a single difference equation. Building on previous work (Deco et al., 2007), the yes-population accumulates noisy evidence for a target, whereas the no-population accumulates a "default" input. Each population has slow recurrent self-excitation (setting the accumulation timescale) and inhibits the other via lateral inhibition. This inherent asymmetry was introduced to account for the asymmetry of the yes-no task, and it was thought to reflect different sizes of the two neural populations in the brain. We set the model parameters so that its RTs (determined by crossing of a decision bound) and fractions of choices qualitatively match the average behaviour of our subjects. We then simulate the impact of different types of pupil-linked neuromodulation – e.g. modulating the gain of all synaptic interactions in the model (Aston-Jones and Cohen, 2005) with a weight that fluctuates from one decision to the next. This work is on-going and will not be part of the paper on this data set.

Pupil-linked brainstem responses

To master the methodological challenges of brainstem fMRI described in section 2.3.1, we have established an approach that follows the recommendations of Eckert, Keren and Aston-Jones (Eckert et al., 2010) and that entails the following steps: We anatomically delineated the LC in each individual by means of melanin-sensitive structural MRI (Shibata et al., 2006) (Figure 70a). We performed fMRI covering the brainstem and large parts of the cortex at a repetition time (TR) of 2 s, with slices oriented perpendicular the longitudinal extent of the LC. To resolve the LC (diameter at least 2 mm) at sufficient signal-to-noise ratio, we chose an in-plane resolution of 2 x 2 mm^2. To further ensure sufficient signal-to-noise ratio, we used a slice thickness of 3 mm, exploiting the fact that the LC spans several centimeters along the longitudinal axis. We monitored cardiac and respiratory cycles concurrently with fMRI and removed their effects from the functional images using an extended version of RETROICOR (Glover et al., 2000). We computed the single-trial task-related fMRI response, separately for each imaging voxel, as the mean across the time window 2 s to 12 s from the start of the decision interval minus the pre-trial baseline activation (-2 s to 2 s from onset of decision interval). We verified that all effects reported below are also evident after excluding short ITIs (data not shown).

Task-related pupil responses (TPR) were robustly coupled to task-related BOLD responses in the individually delineated LC (Figure 70b-e). This coupling was evident when inspecting the fMRI response time courses measured in the LC for high and low TPR bins (Figure 70b, c). These exhibited a regular fMRI response pattern with the typical delay (peak at around 6s from cue onset) for high TPR, and a similar absence of response for low TPR as the corresponding pupil time courses (compare to Figure 68e). The coupling persisted after regressing out residual signal fluctuations in the 4^{th} ventricle after RETROICOR (Figure 70c), as well as other factors such as RT and signal presence. Further, the coupling between TPR and LC responses was not only evident when binning the trials by TPR, but also (and robustly so) at the single trial level (Figure 70d).

Significant coupling between TPR and fMRI responses was also observed for three other brainstem structures: the midbrain nuclei substantia nigra (SN) and ventral tegmental area (VTA) (both coarsely defined based on anatomical atlas coordinates; Figure 70e-g), which release another catecholamine, dopamine and are (directly and indirectly) connected to the LC; and the superior colliculus (SC), which is part of the brain's orienting system and has been linked to pupil dilation in recent invasive work in monkey (Joshi et al., 2016). However, the link to pupil dilation was not evident across all neuromodulatory brainstem centers. For example, the (partly) cholinergic basal forebrain exhibited a task-related deactivation, the amplitude of which was unaffected by TPR (Figure 70g, right panel). A complete map of single-trial TPR-fMRI correlations across the brainstem (cluster-corrected across the brainstem; see red outline in Figure 70e) revealed that, while there was coupling between TPR and neural responses in some regions outside of the LC, this coupling was confined to the structures listed above (Figure 70e). Furthermore, the correlation between TPR and LC-responses was not accounted for by responses in other brainstem regions shown in Figure 70f,g: After removing (via linear regression) their effects, we still obtained a robust TPR-LC correlation in the residuals ($r=0.176$, $p<0.001$). In fact, there was no significant difference in the strength of correlation before and after removal of the influence of the other centers. In sum, while significant coupling to spontaneous trial-to-trial fluctuations in TPR exists for some other brainstem centers, the LC-link is robust and the TPR fluctuations do contain an LC-specific component. Our on-going fMRI work aims at dissociating the pupil-linked response components between the LC and dopaminergic midbrain centers by manipulating, on a single-trial basis, (i) phasic arousal, and (ii) reward prediction errors (by means of external feedback, which was omitted here to allow spontaneous biases to emerge, see above).

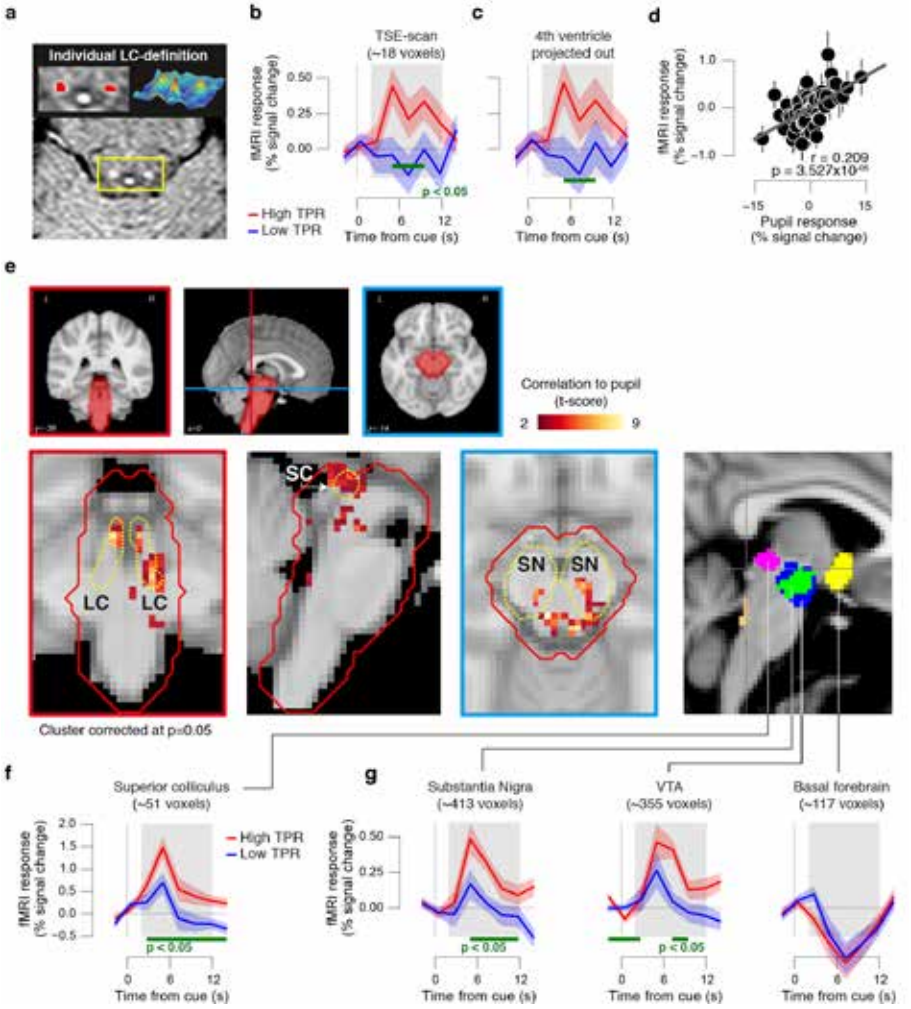


Figure 70: Pupil-linked brainstem responses

(a) Anatomical delineation of the LC in an example subject. In a TSE scan, the LC stands out on each side of the brainstem as a hyper-intense spot at the floor of the 4th ventricle. (b) Task-related BOLD responses in the LC for high and low TPR trials. (c) As in b but after removing (via linear regression) signal fluctuations measured in the 4th ventricle. (d) Single trial correlation between TPR and LC-responses. (e) Cluster-corrected map of single trial correlation between TPR and brainstem responses. (f) As in c but for the superior colliculus (SC). (g) As in c but for substantia nigra (SN), ventral tegmental area (VTA), and basal forebrain (BF). Similar results were obtained by correlating the pre-trial baseline pupil values and fMRI signal levels (data not shown).

Pupil-linked modulation of cortical networks

As commonly observed, the cortical network depicted schematically in Figure 64 activated robustly during the task, while other regions corresponding to the "default mode network" became suppressed (Figure 71a). Task-related fMRI responses across large parts of the cortex were positively coupled to single-trial TPR (Figure 71b), in line with a largely global nature of pupil-linked neuromodulatory influences. However, the spatial correspondence between the maps shown in the top rows of Figures 8a and b were relatively weak (mean correlation = 0.157), and there were several regions that showed significant modulation during task but not by TPR and vice versa (compare top rows between Figure 71a,b).

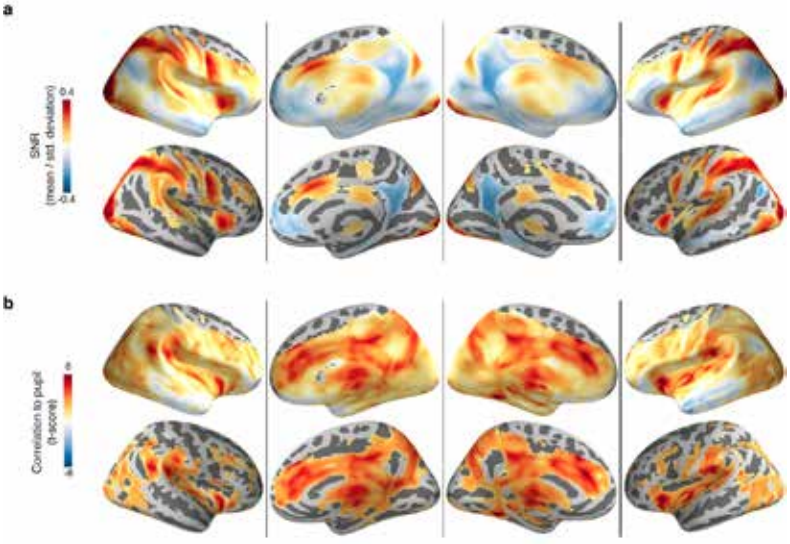


Figure 71: Pupil-linked modulation of task-related cortical responses

(a) Uncorrected (top) and cluster-corrected (bottom) maps of group average SNR of task-related responses (mean divided by standard deviation fMRI responses across trials, computed within each individual). (b) Uncorrected (top) and cluster-corrected (bottom) maps of single trial correlation between TPR and fMRI-response.

Taken together, these results are inconsistent with a simple multiplicative scaling of each voxel's task-related fMRI-response by TPR (as postulated at the single-neuron level for the LC-NA effects, see ref. (Aston-Jones and Cohen, 2005)). They point to a more complex shaping of cortical population activity through phasic arousal. To better characterize and understand that shaping, it will be crucial to compare these maps of pupil-linked metabolic (i.e. BOLD-fMRI) activity changes with the maps of pupil-linked changes of band-limited electrophysiological population activity as assessed by MEG.

In a complementary characterization of the pupil-linked shaping of cortical network dynamics, we parcellated the cortex into 144 regions across both hemispheres. We then computed all-to-all correlations between the trial-to-trial fluctuations of each region's task-related fMRI responses, separately for high and low TPR trials. This revealed that, under high pupil-linked arousal, cortical inter-area correlations were largely reduced. This was quantified by comparing the fraction of region pairs showing an increase with those pairs showing a decrease in correlation under high arousal ($p=0.006$; data not shown). In other words, pupil-linked arousal was accompanied by a predominant suppression of co-fluctuation between the task-related responses of different cortical regions.

Our on-going analyses of the TPR-related changes in cortical responses specifically aim to identify correlates of the TPR-linked reduction in conservative accumulation bias evident in behaviour (Figure 69). To this end, we focus on the two peripheral stages of the sensory-motor decision process: (i) early visual cortical areas encoding the evidence for the contrast target; and (ii) movement-selective areas encoding the action choice. Functional regions of interest (ROIs) in early visual cortex (V1-V3) were defined, within each individual subject, in three steps. First, boundaries between these visual cortical areas were identified by retinotopic mapping via population receptive field imaging (i.e., quantifying, for each voxel, the preferred polar angle and eccentricity). Second, we identified the cortical representation of the stimulus in the yes-no task with a localizer run in each session presenting full-contrast gratings in alternation with blank at the same retinal position as the task-stimulus (for the map of localizer responses in one example subject, see Figure 72a). Third, the ROIs were further constricted by, separately for each

scanning session, taking the 50 voxels that responded (the stimulus "center" sub-region) or deactivated (the stimulus "surround" sub-region) most strongly to the localizer stimulus in the run. Choice-selective regions of interest (ROIs) were identified by mapping significant lateralization with respect to the final hand movement (the mapping between "yes" and "no" perceptual choice and response movement was flipped between both experimental sessions). As expected, we found such movement-selective activity around the hand region of primary motor cortex (M1) in the central sulcus, but also in posterior parietal cortex, in the junction of the intraparietal sulcus and the postcentral sulcus, here referred to as "anterior IPS" (Figure 72g).

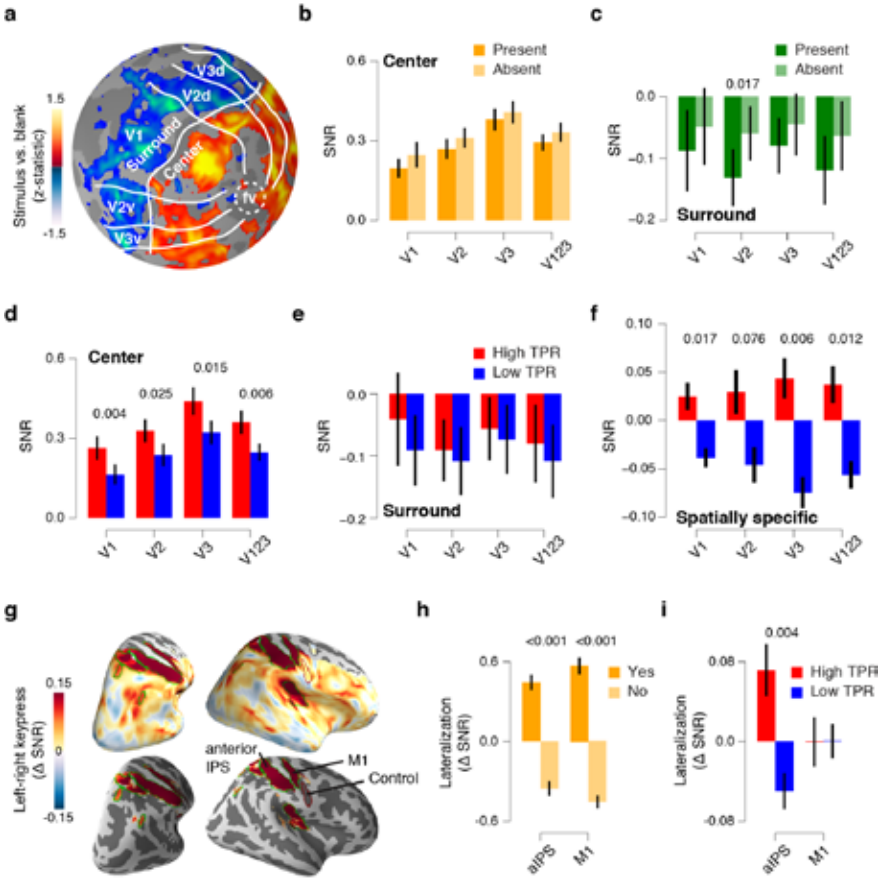


Figure 72: Modulation of selective cortical signals

(a) Flat map of the responses during localizer scans and delineation of areas V1-V3 in an example subject. (b) Task-related fMRI responses in V1-3 center sub-regions, for target present and absent trials. (c) As in b but for surround sub-regions. (d) As in (b), but collapsed across stimulus categories, and separately for high and low TPR. (e) As in d, but for surround sub-regions. (f) As in d, but after removing (via linear regression) single-trial responses from surround, yielding the specific component. (g) Maps of movement-selective lateralization of cortical fMRI responses, displayed on a single hemisphere. (h) Lateralization ("yes-side"- "no-side"), in anterior IPS and M1, for yes and no choice. (i) As in g, but separately for high and low TPR trials.

During the yes-no task, we found positive responses in the center sub-regions of V1-3 (Figure 72b) and suppression in the surround sub-regions (Figure 72c). Different from the stimulus localizer runs, these spatially-specific responses did not reflect the bottom-up stimulus drive: there was no difference between target present vs. absent trials (Figure 72b,c; we do find a robust response to the target stimulus when assessing orientation-selective multi-voxel patterns). We, therefore, interpret the spatially-specific responses shown in Figure 72b-c as a top-down signal in early visual cortex. The magnitude of this top-down signal was modulated by pupil-linked arousal: it was boosted in stimulus sub-regions (Figure 72d), but not in surround sub-regions (Figure 72e). Thus, the TPR-linked

difference was prominent in the spatially-specific component obtained after first removing surround responses from the stimulus sub-region (Figure 72f). Similar TPR-linked differences of the choice-selective activity were observed in anterior IPS, but not in M1 (Figure 72i), although the latter exhibited at least as strong choice-selective activity as the former overall (Figure 72h).

Strikingly, the amount of the modulation of the top-down signal by pupil-linked arousal predicted the amount to which pupil-indexed arousal reduced subjects' conservative decision bias (Figure 73a-c). Again, similar results were obtained in choice-selective regions in posterior parietal cortex (Figure 73d,e) pointing to the interplay between parietal and early visual cortex as a possible mechanism underlying the pupil-linked modulation of behaviour.

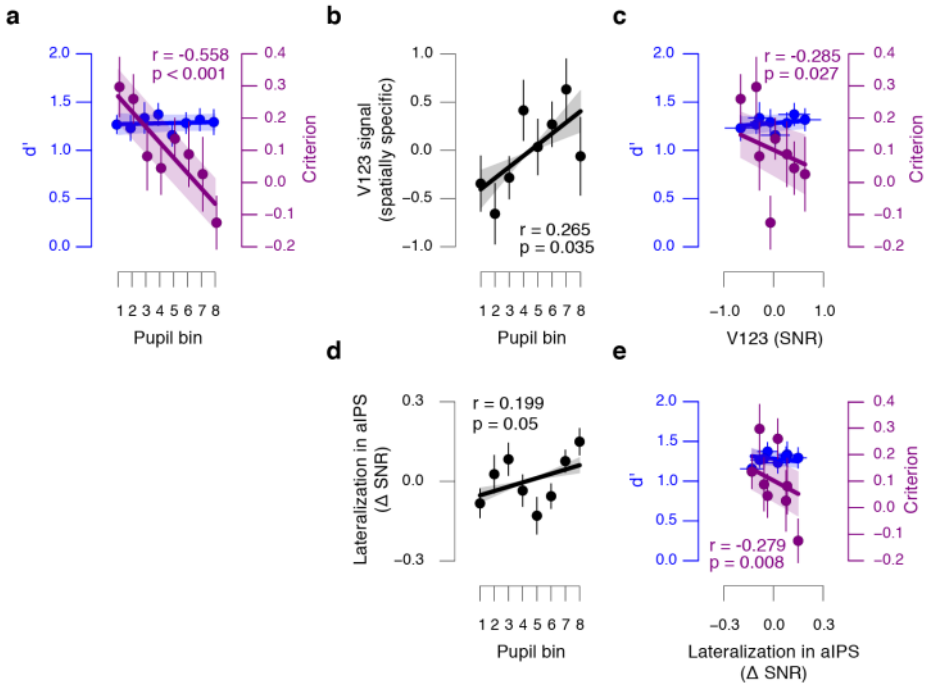


Figure 73: Pupil-linked modulations of selective cortical signals predict behavior

(a) Signal detection theoretic *d'* and *criterion* as a function of TPR (binned). The spatially selective top-down signal in V123 as a function of TPR (binned). (c) Signal detection theoretic *d'* and *criterion* as a function of the selective top-down signal in V123 (both binned by TPR). (d) As in b, but for lateralization in anterior IPS. (e) As in c, but for lateralization in anterior IPS.

Data set 2: Pupil-linked modulation of the cortical dynamics underlying yes vs. no decisions (MEG)

In parallel to the fMRI work reported in the previous *Section*, we collected an MEG data set with this yes-no decision task. The overarching goal here is to study the pupil-linked modulation of the cortical decision *dynamics*, especially the well-characterized dynamical signatures of decision-making reviewed in *Section 2.3.1*. Specific aims were (i) to systematically manipulate decision bias within individuals (rather than relying on spontaneously emerging biases), and (ii) to study the link between pupil-linked arousal state and decision confidence, linking our work to the work reported in *Section 2.1*.

The MEG data will be analysed in the coming months. Our initial analyses have focussed on the behavioural and pupil data, to establish the relationship to the pupil-linked effects in our fMRI experiment, and to look for novel pupil signatures of confidence.

Behavioural task

Participants performed the same yes-no decision-making task as in fMRI (Figure 68a), with the following exceptions: (i) we used two different noise refresh rates (20Hz, and 5Hz), which varied randomly across trials; (ii) two independent online staircase procedures during the main experiment kept each subject's performance at around 75% correct for each noise refresh condition; and (iii) at the end of each trial, we prompted subjects to report how confident they were about their preceding choice and then administered auditory feedback. We had observed that manipulation (i) effectively altered decision criterion (Figure 74a, see below).

Data set

Twenty-three healthy subjects (16 females; age range, 21–31 y) participated in the study. Each subject participated in two scanning sessions to measure MEG responses in the main experiment (two hours per session).

Pupil-linked modulation of behaviour

The noise refresh manipulation strongly affect subjects' decision bias: they were slightly liberal in the fast noise refresh condition, but very conservative in the slow noise refresh condition (Figure 74a). Importantly, we found the same pattern of pupil-linked behavioural modulation as in the fMRI data set: a conservative bias (here computed as signal detection theory *criterion*) was reduced on high TPR trials, while sensitivity (*d'*) was unaffected (Figure 74a). Furthermore, we found that TPR also inversely scaled with subjects' decision confidence, with the strongest dilations occurring on the very unsure trials (Figure 74b). Our planned analyses will of the MEG data will specifically screen for signatures predicting these clear, pupil-linked behavioural effects.

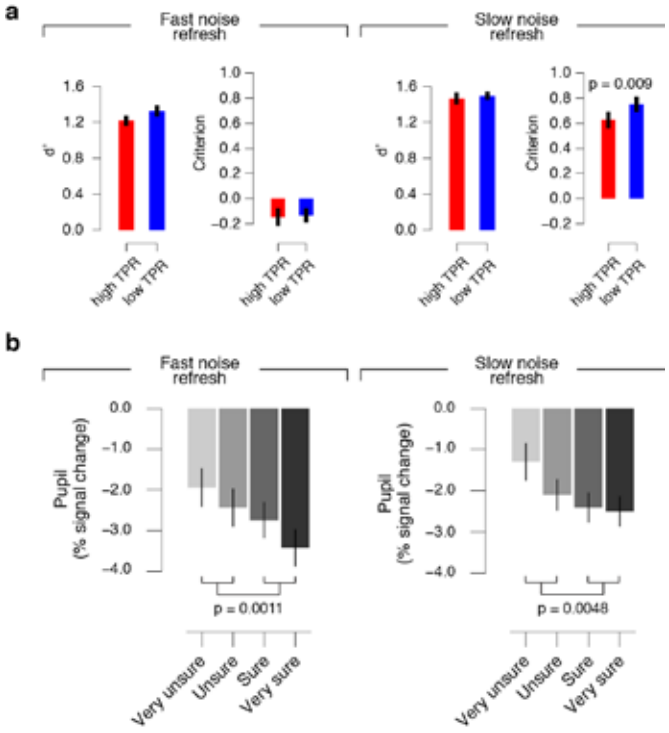


Figure 74: Pupil-linked modulation of behaviour

(a) Signal detection theory measures *d'* and *criterion* separately for each noise refresh condition, and separately for high and low TPR trials. (b) Decision related pupil responses scale with decision-confidence.

Two **Dataset Information Cards** have been completed (see DICs Task T3.2.3 "Brainstem modulation of decision processes (human behaviour and MEG)").

Provenance of data and location of data storage

We used the funding from the HBP to collect 2 neuroimaging data sets: an fMRI data set (at UvA) and an MEG data set (at UKE). The fMRI dataset was collected by Jan Willem de Gee (UKE) and Olympia Colizoli (UvA). The MEG dataset was collected by Jan Willem de Gee (UKE) and Niels Kloosterman (UvA). These data sets are described in detail in the subsequent sections. In addition, we analyzed several other data sets (collected from other funds) to characterize the neuromodulation of cortical decision processing.

The HBP-funded data sets are located at the following links:

- Data set #1: http://s3.data.kit.edu/SP3/3_2_3/Study1_yesno_fMRI
- Data set #2: http://s3.data.kit.edu/SP3/3_2_3/Study2_yesno_MEG

Self-analysis of the value and completeness of data

The *Data Sets 1* and *2* are complete and were deposited on the HBP server at the above links. The analysis and interpretation is close to complete for *Data Set 1* and on-going for *Data Set 2*. Further, the (re-)analysis of several other data sets, collected outside of HBP, has provided additional, complementary insights into the modulation of cortical decision processing (see publications below). Finally, Tobias Donner is working on a review on the state-dependent modulation of cortical decision processing that integrates results from recent work conducted in the rodent, monkey, and human brain, including the results from the HBP *Task 3.2* Taken together, our work within HBP has filled an important gap in the

literature on decision-making.

Publications, disseminations, collaborations

a. Peer-reviewed publications connected to HBP work (HBP acknowledged):

- Tsetsos K, Pfeffer T, Jentgens P & Donner TH. 2015. Action Planning and the Timescale of Evidence Accumulation. *PLoS One* 12;10(6): e0129473. doi: 10.1371/journal.pone.0129473. (Raw data available at https://osf.io/ju95a/)
- Kloosterman NA, Meindertsma T, Hillebrand A, van Dijk B, Lamme VAF & Donner TH. 2015. Top-Down Modulation in Human Visual Cortex Predicts the Stability of a Perceptual Illusion. *Journal of Neurophysiology* 113: 1063-76.

b. Conference abstracts on HBP Data Sets and results connected to HBP work (HBP acknowledged):

- De Gee JW, Kloosterman NA, Nieuwenhuis S, Knapen T & Donner TH. 2015. Decision-related pupil dilation reflects locus coeruleus activity and altered visual evidence accumulation. No. 2015-S-6760-SfN. *Chicago, IL: Society for Neuroscience*.
- Kloosterman NA, de Gee JW, & Donner TH. Effects of noradrenaline on visual evidence accumulation in human cortex. No. 2015-S-11496-SfN. *Chicago, IL: Society for Neuroscience*.
- Meindertsma T, Kloosterman NA, Nolte G, Engel AK, & Donner TH. 2015. Decision-related oscillatory activity in human visual cortex is linked to pupil dilation. No. 2015-S-9003_SfN. *Chicago, IL: Society for Neuroscience*.
- Urai A, De Gee JW, & Donner TH. Eye opener: Pupil dilation signals decision uncertainty. No. 2015-S-2663-SfN. *Chicago, IL: Society for Neuroscience*.

c. Manuscripts on HBP work in preparation:

- Manuscript on *Data set #1* (Working title: *Pupil-linked arousal systems shape cortical state and choice behavior*).
- Integrative review (Working title: *State-dependent modulation of perceptual decision-making*).
- Manuscript on HBP-connected work (Working title: *Pupil-linked modulation of decision-related top-down signal in visual cortex*).
- Manuscript on HBP-connected work (Working title: *Decision uncertainty drives pupil-linked arousal systems and modulates sequential choice bias*).

d. Invited platform presentations by Tobias Donner on HBP-connected work (HBP acknowledged):

- Cosyne workshop *Form and function of Choice-related Feedback Signals in Decision Making,* Snowbird. 1 March 2016.
- European Institute of Theoretical Neuroscience (EITN) workshop *Probabilistic Inference and the Brain*, Paris. 10 September 2015.
- Dutch Neuroscience Meeting symposium *Perceptual Decision-making*. Lunteren. 12 June 2015.
- ICPS symposium *Model-based Neuroscience of Strategic Decision-making*, Amsterdam. 13 March 2015.
- Tübingen MEG Symposium. 27 October 2014.
- Symposium *The Many Faces of Top-Down: an Integrative Perspective*. Organization for Human Brain Mapping, Hamburg. 10 June 2014. (Chair and Speaker)
- Donders Institute symposium *Dialogues on the Role of Top-down Factors in Sensory Processing*. Radboud University Nijmegen. 21 May 2014.

e. Collaborations on this topic within HBP:

- Florent Meyniel, Mariano Sigman, Stanislas Dehaene (SP3 T 3.1.1, ongoing): Neural bases of decision confidence and uncertainty.
- Gustavo Deco (SP4 T4.3.1, ongoing): Large-scale modeling of state-dependent modulation of cortical networks.
- Cyriel Pennartz (new SP3, ongoing): Neuromodulation of cortical decision dynamics.
- Mathias Pessiglione (SP3 T3.2.2, planned): Cost-benefit analysis of perceptual evidence accumulation.
- Avi Karni (SP3 T3.3.1, planned): Pupillometric correlates of motor skill learning.

f. Collaborations on this topic outside of HBP:

- Marius Usher (Tel Aviv University): Modeling the neuromodulation of evidence accumulation and decision-making.
- David McCormick, Matthew McGinley (Yale University): Comparison of pupil-linked modulation of perceptual decision-making in humans and mice.

2.4 Characterizing the brain architecture of decision-related motivational states and values

Task T3.2.4 - Talma Hendler (TASMC), Itzhak Fried (TASMC), Tomer Gazit (TASMC)

Overview

Perspective on motivational decision-making

What is motivation?

As living organisms we all engage in day-to-day environmental challenges: we need food, water and sleep in order to survive; as social beings we also need love and appreciation in order to maintain our emotional wellbeing, and as part of the modern society we need money and estate in order to provide ourselves with a roof, clothes and practically every material need we have. At the same time we have to be aware of dangers, which may be actually life-threatening, or carry the potential for physical and/or emotional pain. Since achieving these needs and escaping danger is essential for our survival, environmental stimuli which meet these needs are experienced as rewarding and appetitive, and we are driven to approach them, whereas we are driven to avoid the dangerous and painful stimuli which are experienced as punishing and aversive. These reinforcing accounts and behavioral drives are generated by the psycho-behavioral process of motivation, which is one of our basic survival mechanisms.

Motivational processes start with the detection and assessment of environmental incentives and threats in order to facilitate goal directed behavior that promotes survival and wellbeing of the organism. Thus, motivational processes can be viewed as a key behavioral modulator, mediating interactions with the environment and adaptive behavior. Importantly, in the real world these interactions are of a dynamic, progressive nature, which require complex yet rapid computations regarding optional stimulus-response scenarios and effective behavioral decision making. Inappropriate choices of motivational behavior may lead to overt psychopathology, such as generalized behavioral inhibition and avoidance in anxiety (Hendler et al., 2014; McNaughton and Corr, 2008) and excessive goal directed behavior in the manic state of bipolar disorder (Gonen et al., 2014; Johnson et al., 2012). Thus, neural characterization of motivational processes as well as the interactions between them, may lead to better understanding of pathological behaviors.

However, motivationally driven behavior is determined following multidimensional computations encompassing different features of the reinforcement: its different accounts (incentive and hedonic) and the expected outcome (reward or punishment). Thus, motivational processing is based on the ongoing assessment of various internal and external cues according to their objective and subjective value, generating complex estimations aimed at producing the most adaptive behavior (approach or avoidance) at any given moment. Different aspects of this multilayered process, such as the neurochemistry (Berridge and Kringelbach, 2013a) and mechanistic circuitry (Haber and Knutson, 2009) of the reward circuit, reward learning (Dayan and Berridge, 2014; Pizzagalli et al., 2008), or effort-based decision making (Salamone and Correa, 2012), have been widely investigated. Nevertheless, deconstruction of the motivational mechanism into its basic elements and stages in order to unveil the way they interact and are regulated in the human brain in order to promote adaptive behavioral decision-making is still lacking.

The current review aims to provide a general overview deconstructing motivation into states (reward, punishment and goal conflict), accounts (incentive, hedonic) and behavior (approach, avoidance); providing a glance on their neural correlates as well. Further, as motivational tendencies have been related to individual differences and personality (Corr and McNaughton, 2012; Smillie, 2008), we characterize these relationships at both the behavioral and neural levels.

Motivational states

The idea that individuals' behavior may be modulated by their sensitivity to different kinds of motivational states originated at the beginning of the 20th century with the seminal work of Ivan Pavlov who discovered the striking effect of reward and punishment on behavior (Corr and Perkins, 2006). This groundbreaking work laid the foundation for a later motivation theory pointing to the role of sensitivity to potential rewards and punishments in the organism's tendency to act upon or away from a stimulus in the environment (i.e. "Reinforcement Sensitivity Theory": RST; (Smillie, 2008). Based on animal research (Gray, 2000; McNaughton, 2008), a neurobehavioral model including three distinct subsystems involved in motivational processes was suggested: 1) The "Fight Flight Freeze System" (FFFS), mediates sensitivity to punishment (and non-reward) via noradrenergic functions (McNaughton and Corr, 2004). 2) The "Behavioral Activation System" (BAS), underlies sensitivity to reward (and non-punishment) via dopaminergic pathways (Depue and Collins, 1999; Smillie, 2008). 3) The "Behavioral Inhibition System" (BIS), sensitive to goal-conflict situations, is thus activated by stimuli of mixed or ambiguous values (i.e. reward and punishment) and is mediated by serotonin function (Gray and McNaughton, 2000). The system serves to compare between the current state, previous knowledge and expected consequences which are all used for adaptive behavioral selection. The RST model provides a comprehensive scheme for three separate neural sub-systems of motivation, describing their relation to behavioral interaction with the environment (for the neural representation of the three systems see Figure 75).

While the RST has been established in animal studies, evidence regarding the human brain is less comprehensive. We have recently found, using an interactive “Domino” game, that different motivational states elicited activations in brain regions that corresponded to the brain systems underlying RST. Moreover, using Dynamic Causal Modeling (DCM) for each motivational system, we confirmed that the coupling strengths between the key brain regions of each system were enabled selectively by the appropriate motivational state(Gonen et al., 2012).

Importantly, this conceptualization is insufficient for capturing the full complexity of motivational processes and their manifestation in the healthy and pathological forms, since every motivational state – reward or punishment - carries two distinct accounts – incentive and hedonic, which must be addressed when considering regulation of motivational drives.

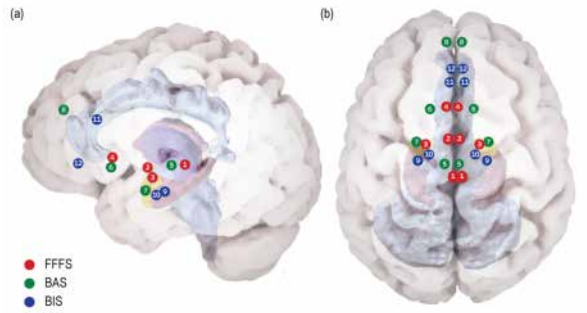


Figure 75: Neuroanatomy of the RST motivational systems.

According to RST three bio-behavioral systems participate in reinforcement modulation of goal directed behavior: 1. The **Fight Flight Freeze System (FFFS)** is activated by all punishment stimuli (shown in red). This system includes the PAG (1), Medial Hypothalamus (2), central Amygdala (3) and sgACC (4). **2. The Behavioral Activation System (BAS)** believed to underlie reward (and non-punishment) sensitivity (shown in green). The system relies on VTA dopamine phasic activity (5) to NAC (6) in response to reward. Information regarding integrative stimulus-reward associations is projected to the NAC from the basolateral amygdala (7). The dorso-medial pre-frontal cortex carries integrative representation of complex reinforcement associations with both stimuli and responses (8). 3. **Behavioral Inhibition System (BIS)** underlying goal-conflict situations (shown in blue), consists of two neural foci: the Septo-Hippocampal System (SHS) (9) which is informed comprehensively regarding possible behavioral plans for the current situation and their consequences by the entorhinal cortex (10) and ACC (11). The ventro-medial pre-frontal cortex (12) is considered a behavioral control modulator. (This figure is adapted from (Gonen et al., 2014)).

While the RST has been established in animal studies, evidence regarding the human brain is less comprehensive. We have recently found, using an interactive "Domino" game, that different motivational states elicited activations in brain regions that corresponded to the brain systems underlying RST. Moreover, using Dynamic Causal Modeling (DCM) for each motivational system, we confirmed that the coupling strengths between the key brain regions of each system were enabled selectively by the appropriate motivational state(Gonen et al., 2012).

Importantly, this conceptualization is insufficient for capturing the full complexity of motivational processes and their manifestation in the healthy and pathological forms, since every motivational state – reward or punishment - carries two distinct accounts – incentive and hedonic, which must be addressed when considering regulation of motivational drives.

Motivational accounts

Reinforcements in the environment signal incentive accounts of goal-directed information, according to their expected effect, either reward or punishment; along with hedonic accounts holding the affective information according to the feeling they evoke, either appetitive or aversive (Berridge and Kringelbach, 2008). The combination of these two accounts guides the preferable motivational behavioral choice: approach or avoidance. In everyday life, incentive and hedonic accounts interact in either a congruent or incongruent way. For example, congruence occurs when one enjoys playing basketball, which is also good for one's health (hedonic and incentive accounts are both positive), or dislikes butter which is bad for one's health (hedonic and incentive accounts are both negative). Incongruence on the other hand occurs for example in the case of eating a delicious cake when on a diet (the hedonic account of eating the cake is positive but the incentive account is negative), or when your friends call you for an afternoon drink but you need to stay at work on an important project (the hedonic account of staying at work is negative but the incentive account is positive – it will serve your career).

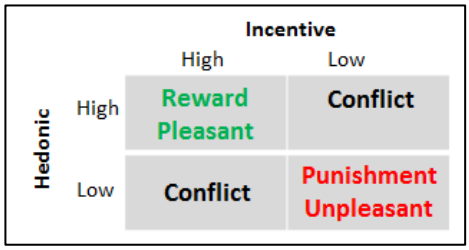


Figure 76: Demonstration of possible relationships between the incentive and hedonic motivational accounts.

These contradicting situations raise an internal conflict in terms of what promotes one's goals as opposed to what is pleasant (Berridge, 2009), a conflict which has to be resolved in order to choose the most adaptive behavior (i.e. approach or avoidance). Maladaptive behavioral choices may have tolerable consequences when they happen infrequently, yet when it becomes common practice it could lead to abnormal conditions, such as excessive drug consumption in addictions (Diekhof et al., 2008; Reuter et al., 2005) or starvation in eating disorders (Berridge, 1996, 2009).

The generation and modulation of these two attributes are known to rely on different neurochemical processes: dopamine has been assumed to mediate the incentive component via the mesolimbic pathway (ventral tegmental area [VTA] -nucleus accumbence [NAC] – medial prefrontal cortex [PFC]) and facilitate behavior, while opioid neurotransmission within the ventral striatum, and especially in the NAC is suspected to indicate hedonic value (Berridge and Kringelbach, 2013b; Berridge and Robinson, 2003). In the lab the incentive value, (also referred to in the literature as 'wanting' (Berridge, 2009) or 'motivational salience' (Roesch and Olson, 2004)), is commonly manipulated by reinforcement novelty or magnitude; while the hedonic value, (also referred to in the literature as 'liking' (Berridge and Kringelbach, 2008)), is commonly manipulated by the valence of the reinforcement (appetitive or aversive). There have been several efforts to dissociate these processes at the neural level in both animals and humans. Roesch& Olson (Roesch and Olson, 2004) manipulated both reinforcement valence (hedonic

value) and magnitude (incentive value) while recording from the left lateral orbito-frontal cortex (OFC) and the premotor (PM) regions of the macaque brain. They concluded that neurons in the OFC selectively represent the hedonic value and those in the PM respond to the incentive value. Others have identified the NAC as sensitive to the hedonic value of reinforcement while the baso-lateral amygdala serves to enhance its incentive value (Chang et al., 2012). However, recent studies have suggested that both incentive and hedonic value are represented by distinct neural populations within the NAC (Bissonette et al., 2013; Peciña et al., 2006). Several studies have tackled this issue in humans, using functional neuroimaging. Anderson and colleagues (Anderson et al., 2003) used pleasant and unpleasant odors in different concentrations to manipulate both valance (hedonic) and intensity (incentive). They specifically looked at the amygdala and OFC, and found that the activity in the amygdala was related to stimulus intensity regardless of its valence and activity in the medial OFC was related to valence regardless of intensity. Others have shown sensitivity to intensity across valences in the amygdala as well as in other salience-related regions (e.g. pons and insula), while preference to valence irrespective of intensity was found in the lateral OFC but also in other paralimbic areas (e.g. Anterior Cingulate Cortex, ACC)(Small et al., 2003). Of note is that other regions have been indicated in investigations focused on deciphering only one of these processes. For example, PCC was found related to hedonic value and not incentive (Litt et al., 2011), while Supplementary Motor Area (SMA) premotor and primary motor cortices, as well as the thalamus, were previously related to incentive and not hedonic (e.g. (Huettel et al., 2005; Pessiglione et al., 2007). Nonetheless, like in the animal studies, mixed findings exist in humans as well. For example, some had linked the OFC to salience coding, irrespective of valance (e.g.(Diekhof et al., 2011)). Similarly, with regard to the NAC, some have argued for its involvement in incentive salience coding (e.g. (Jensen et al., 2007; Zink et al., 2006)) and not only hedonic (Gottfried, O'Doherty et al. 2002; Haber and Knutson 2009).

Motivational behavior

Within the incentive process of motivation lays the most important output of motivational decision making: the resulting behavior. Human studies of motivation have mostly used paradigms with static stimuli, presented in a trial-by-trial manner, providing the opportunity to separately investigate different stages of the motivational brain response, such as the anticipation, outcome, and evaluation of rewards and punishments (e.g. (Gonen et al., 2012; Liu et al., 2011). Several studies have used manipulations of reinforcement saliency, magnitude or intensity as operationalization of the incentive account of motivation, looking at the anticipation or response to reinforcement (Huettel et al., 2005; Pessiglione et al., 2007). In recent years neuroimaging studies of motivation have elucidated several stages in this complex stimulus-response scenario. Findings regarding the initial valuation and processing of reinforcements point first and foremost to the relevance of the dopaminergic mesostriatal pathway, including the Ventral Striatum (VS) and Ventral Tegmental Area (VTA) (Haber and Knutson, 2009), in both animals (Cardinal *et al.*, 2002) and humans (Liu *et al.*, 2011). Accordingly, a recent fMRI study demonstrated that the facilitating effects of incentive motivation involved the caudate and putamen (Miller *et al.*, 2014). Nonetheless, despite our growing knowledge of motivational processing, including cost-benefit valuation during behavioral decision-making (Basten *et al.*, 2010; Park *et al.*, 2011), surprisingly few studies examined the neural processes underlying the behavioral phase of incentive motivation: approach or avoidance (Bach et al., 2014). The paucity of neuroimaging studies investigating motivational behavior could result from the conventional operationalization used in functional-imaging, of static stimuli presented in a trial-by-trial manner (c.f. (Liu et al., 2011); which does not allow for characterization of dynamic ongoing motivational behavior. Such behaviors should be investigated using ecological tools which will simulate the rapidly unfolding nature of such a dynamic process; as we have recently demonstrated (Gonen et al., 2016).

Motivation and individual differences

Motivational behavioral tendencies have been long associated with individual differences in personality. Three main personality models have been related to the tendency to approach or avoid environmental incentives and threats. First, the phenomenologically driven model NEO-five-factor inventory (NEO-FFI) (Costa and McCrae, 1992) is one of the most prominent models for personality structure, consisting of five broad traits of personality. Of these, two traits were repeatedly related to motivational tendencies: Extraversion (E) and Neuroticism (N), previously shown to be related to the tendency to approach and avoid, respectively (Canli, 2004; Canli et al., 2002; Cohen et al., 2005). Another approach to personality structure, driven by a neurochemical mechanistic perspective, was suggested by Cloninger (Cloninger, 1987) positing the "Tridimensional model", including the traits Novelty Seeking (NS), and Reward Dependence (RD), which were related to approach tendencies, and Harm Avoidance (HA), related to avoidance tendency (Cloninger, 1987). Finally, the RST maintains that it is individual sensitivity to reinforcement which guides behavior. Within the framework of RST, an independent measure was developed for the traits sensitivity to reward (SR) and sensitivity to punishment (SP), which were suggested to underlie approach and avoidance tendencies, respectively (Torrubia, 2001). Although each of these personality models have been previously related to motivational tendencies, in both the behavioral and neural levels (for review see (Smillie, 2008); accumulating evidence are not conclusive and moreover, to only few study has investigated the neural correlates of these traits while actually performing a dynamic motivational task. We have recently used an original integrative profiling of individual differences derived from three theoretical models, and showed that high scores in approach related traits (extraversion, reward-dependence and agreeableness) were manifested in increased tendency to approach behavior during the game, mostly under high goal-conflict. Furthermore, decreased tendency to approach was manifested by high scores in avoidance

related traits (harm-avoidance, neuroticism and punishment-sensitivity) (Gonen et al., 2016). Interestingly, during high goal conflict, the approach group had greater activation in both VTA and VS compared to the avoidance group, while no difference was found between the groups in the activity of other brain-regions. With relation to individual tendencies, this finding provides a neural distinction between bottom-up processing in the VTA and VS and top-down processing in the medial prefrontal region. The significance of bottom-up processing to individual differences further implies that motivational tendencies and their neural correlates reflect the innate temperamental parts of personality structure, compatible with Cloninger's tridimensional model (Cloninger, 1987). Yet it was only the integrative personality profile we constructed that demonstrated such multilevel relations of personality and neurobehavioral patterns, indicating that the tridimensional model itself could not account for individual differences during actual motivational behavior. Pointing to bottom-up temperamental parts of personality rather than the acquired top-down influences is an important distinction: abnormal top-down regulatory mechanisms (such as the PFC) have been suggested to underlie motivation related psychopathological conditions (Choi et al., 2011; Phillips, 2008); yet our data suggest that the incentive related, low level facilitators of the dopaminergic mesostriatal pathway (such as the VS or VTA) may serve as stronger candidates for variations in motivational behavior tendencies.

REFERENCES

Anderson, A.K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D., Glover, G., Gabrieli, J., and Sobel, N. (2003). Dissociated neural representations of intensity and valence in human olfaction. Nature Neuroscience 6, 196-202.

Bach, D.R., Guitart-Masip, M., Packard, P.A., Miró, J., Falip, M., Fuentemilla, L., and Dolan, R.J. (2014). Human Hippocampus Arbitrates Approach-Avoidance Conflict. Current Biology 24, 541-547.

Basten, U., Biele, G., Heekeren, H.R., and Fiebach, C.J. (2010). How the brain integrates costs and benefits during decision making. Proceedings of the National Academy of Sciences 107, 21767-21772.

Berridge, K.C. (1996). Food reward: brain substrates of wanting and liking. Neuroscience & Biobehavioral Reviews 20, 1-25.

Berridge, K.C. (2009). 'Liking'and 'wanting'food rewards: Brain substrates and roles in eating disorders. Physiology & Behavior 97, 537-550.

Berridge, K.C., and Kringelbach, M.L. (2008). Affective neuroscience of pleasure: reward in humans and animals. Psychopharmacology 199, 457-480.

Berridge, K.C., and Kringelbach, M.L. (2013a). Neuroscience of affect: brain mechanisms of pleasure and displeasure. Current Opinion In Neurobiology 23, 294-303.

Berridge, K.C., and Kringelbach, M.L. (2013b). Neuroscience of affect: brain mechanisms of pleasure and displeasure. Current Opinion in Neurobiology.

Berridge, K.C., and Robinson, T.E. (2003). Parsing reward. Trends in neurosciences 26, 507-513.

Bissonette, G.B., Burton, A.C., Gentry, R.N., Goldstein, B.L., Hearn, T.N., Barnett, B.R., Kashtelyan, V., and Roesch, M.R. (2013). Separate Populations of Neurons in Ventral Striatum Encode Value and Motivation. Plos One 8, e64673.

Canli, T. (2004). Functional brain mapping of extraversion and neuroticism: Learning from individual differences in emotion processing. Journal of Personality 72, 1105.

Canli, T., Sivers, H., Whitfield, S.L., Gotlib, I.H., and Gabrieli, J.D.E. (2002). Amygdala response to happy faces as a function of extraversion. Science 296, 2191.

Cardinal, R.N., Parkinson, J.A., Hall, J., and Everitt, B.J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neuroscience & Biobehavioral Reviews 26, 321.

Chang, S.E., Wheeler, D.S., and Holland, P.C. (2012). Roles of nucleus accumbens and basolateral amygdala in autoshaped lever pressing. Neurobiology of learning and memory 97, 441-451.

Choi, S.W., Chi, S.E., Chung, S.Y., Kim, J.W., Ahn, C.Y., and Kim, H.T. (2011). Is Alpha Wave Neurofeedback Effective with Randomized Clinical Trials in Depression? A Pilot Study. Neuropsychobiology 63, 43-51.

Cloninger, C. (1987). A systematic method for clinical description and classification of personality variants : A proposal. Archives of General Psychiatry 44, 573.

Cohen, M.X., Young, J., Baek, J.-M., Kessler, C., and Ranganath, C. (2005). Individual differences in extraversion and dopamine genetics predict neural reward responses. Cognitive Brain Research 25, 851.

Corr, P.J., and McNaughton, N. (2012). Neuroscience and approach/avoidance personality traits: A two stage (valuation-motivation) approach. Neuroscience & Biobehavioral Reviews.

Corr, P.J., and Perkins, A.M. (2006). The role of theory in the psychophysiology of personality: From Ivan Pavlov to Jeffrey Gray. International Journal of Psychophysiology 62, 367.

Costa, J.P.T., and McCrae, R.R. (1992). Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory. Psychological Assessment 4, 5.

Dayan, P., and Berridge, K.C. (2014). Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation. Cognitive, Affective, & Behavioral Neuroscience, 1-20.

Depue, R.A., and Collins, P.F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. Behavioral and Brain Sciences 22, 491.

Diekhof, E.K., Falkai, P., and Gruber, O. (2008). Functional neuroimaging of reward processing and decision-making: A review of aberrant motivational and affective processing in addiction and mood disorders. Brain Research Reviews 59, 164.

Diekhof, E.K., Falkai, P., and Gruber, O. (2011). The orbitofrontal cortex and its role in the assignment of behavioural significance. Neuropsychologia 49, 984-991.

Gonen, T., Admon, R., Podlipsky, I., and Hendler, T. (2012). From Animal Model to Human Brain Networking: Dynamic Causal Modeling of Motivational Systems. The Journal of Neuroscience 32, 7218-7224.

Gonen, T., Sharon, H., Pearlson, G., and Hendler, T. (2014). Moods as ups and downs of the motivation pendulum: Revisiting Reinforcement Sensitivity Theory (RST) in Bipolar Disorder. Name: Frontiers in Behavioral Neuroscience 8, 378.

Gonen, T.S., Eyal , Eldar, E., Ben Simon, E., Raz, G., and Hendler, T. (2016). Human mesostriatal response tracks motivational

tendencies under naturalistic goal-conflict. Social Cognitive Affective Neuroscience.

Gray, J.A., and McNaughton, N. (2000). The Neuropsychology of Anxiety: an Enquiry in to the Functions of the Septo-hippocampal System, 2nd Edition edn (Oxford: Oxford University Press).

Gray, J.A., McNaughton, N. (2000). The Neuropsychology of Anxiety: an Enquiry in to the Functions of the Septo-hippocampal System, 2nd Edition edn (Oxford: Oxford University Press).

Haber, S.N., and Knutson, B. (2009). The Reward Circuit: Linking Primate Anatomy and Human Imaging. Neuropsychopharmacology 35, 4-26.

Hendler, T., Gonen, T., Harel, E.V., and Sharon, H. (2014). From Circuit Activity to Network Connectivity and Back: The Case of Obsessive-Compulsive Disorder. Biological Psychiatry 75, 590-592.

Huettel, S.A., Song, A.W., and McCarthy, G. (2005). Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. The Journal of Neuroscience 25, 3304-3311.

Jensen, J., Smith, A.J., Willeit, M., Crawley, A.P., Mikulis, D.J., Vitcu, I., and Kapur, S. (2007). Separate brain regions code for salience vs. valence during reward prediction in humans. Human Brain Mapping 28, 294-302.

Johnson, S.L., Edge, M.D., Holmes, M.K., and Carver, C.S. (2012). The behavioral activation system and mania. Annual review of clinical psychology 8, 243.

Litt, A., Plassmann, H., Shiv, B., and Rangel, A. (2011). Dissociating valuation and saliency signals during decision-making. Cerebral Cortex 21, 95-102.

Liu, X., Hairston, J., Schrier, M., and Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. Neuroscience & Biobehavioral Reviews 35, 1219-1236.

McNaughton, N., and Corr, P.J. (2004). A two-dimensional neuropsychology of defense: Fear/anxiety and defensive distance. Neuroscience & Biobehavioral Reviews 28, 285.

McNaughton, N., and Corr, P.J. (2008). The neuropsychology of fear and anxiety: A foundation for reinforcement sensitivity theory. In The Reinforcement Sensitivity Theory of Personality, P.J. Corr, ed. (Cambridge: Cambridge University Press), pp. 44-94.

McNaughton, N., Corr, P.J. (2008). The neuropsychology of fear and anxiety: A foundation for reinforcement sensitivity theory. In The Reinforcement Sensitivity Theory of Personality, P.J. Corr, ed. (Cambridge: Cambridge University Press), pp. 44-94.

Miller, E.M., Shankar, M.U., Knutson, B., and McClure, S.M. (2014). Dissociating Motivation from Reward in Human Striatal Activity. Journal of Cognitive Neuroscience 26, 1075-1084

Park, S.Q., Kahnt, T., Rieskamp, J., and Heekeren, H.R. (2011). Neurobiology of value integration: when value impacts valuation. The Journal of Neuroscience 31, 9307-9314.

Peciña, S., Smith, K.S., and Berridge, K.C. (2006). Hedonic hot spots in the brain. The Neuroscientist 12, 500-511.

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R.J., and Frith, C.D. (2007). How the brain translates money into force: a neuroimaging study of subliminal motivation. Science 316, 904-906.

Phillips, M. (2008). A neural model of voluntary and automatic emotion regulation : Implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. Molecular Psychiatry 13, 833-857.

Pizzagalli, D.A., Goetz, E., Ostacher, M., Iosifescu, D.V., and Perlis, R.H. (2008). Euthymic patients with bipolar disorder show decreased reward learning in a probabilistic reward task. Biological Psychiatry 64, 162-168.

Reuter, J., Raedler, T., Rose, M., Hand, I., Gläscher, J., and Büchel, C. (2005). Pathological gambling is linked to reduced activation of the mesolimbic reward system. Nature Neuroscience 8, 147-148.

Roesch, M.R., and Olson, C.R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. Science 304, 307-310.

Salamone, J.D., and Correa, M. (2012). The mysterious motivational functions of mesolimbic dopamine. Neuron 76, 470-485.

Small, D.M., Gregory, M.D., Mak, Y.E., Gitelman, D., Mesulam, M., and Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. Neuron 39, 701-711.

Smillie, L.D. (2008). What is reinforcement sensitivity? Neuroscience paradigms for approach-avoidance process theories of personality. European Journal of Personality 22, 359.

Torrubia, R., Ávila, C., Moltó, J., Caseras, X. (2001). The sensitivity to punishment and sensitivity to reward questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. Personality and Individual Differences 3, 837–862.

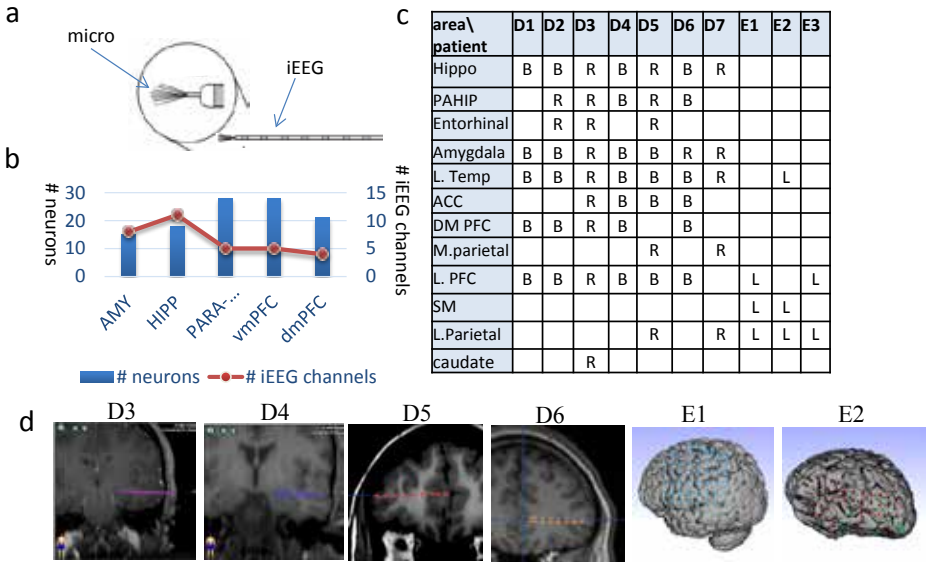
Zink, C.F., Pagnoni, G., Chappelow, J., Martin-Skurski, M., and Berns, G.S. (2006). Human striatal activation reflects degree of stimulus saliency. NeuroImage 29, 977-983.

Data set: Intracranial single cell and LFP dataset

Introduction

Motivation is a key aspect in human behaviour and decision making; shaped by the interaction between environmental reinforcing cues and the organism's specific goals. Reinforcements signal both incentive (reward/punishment) and hedonic (appetitive/aversive) accounts, with their integration often raising a conflict considering the preferable action choice (approach/avoidance). Pathological resolution of such conflicts is often evident in psychiatric symptoms, such as diminished or excessive approach in depression and addiction (respectively), and avoidance in eating disorders. Depicting the neural signature of motivational dimensions could thus advance both domain based brain diagnosis and treatment (RDoC).

In this project we aimed to illuminate the multilevel neural architecture that underlies motivational decision making by deconstructing it into stages: appearance of the incentive cue, decision on action under conflict (to approach or avoid), action (approaching or avoiding), anticipation to outcome and response to outcome (reward or punishment).



c

area\ patient	D1	D2	D3	D4	D5	D6	D7	E1	E2	E3
Hippo	B	B	R	B	R	B	R			
PAHIP		R	R	B	R	B				
Entorhinal		R	R		R					
Amygdala	B	B	R	B	B	R	R			
L. Temp	B	B	R	B	B	B	R		L	
ACC			R	B	B	B				
DM PFC	B	B	R	B		B				
M.parietal					R		R			
L. PFC	B	B	R	B	B	B		L		L
SM								L	L	
L.Parietal					R		R	L	L	L
caudate			R							

Figure 77: Recording sites

(a) Image of the Benkhe Fried electrodes used for depth recording. (b) Quantification of the number of neurons recorded (blue) and the number of iEEG channels for the 5 medial regions elaborated in this report. (c) Table summarizing the data collected from different brain regions (PAHIP= parahippocampus, L Temp=lateral temporal, ACC=anterior cingulate, M. Parietal=medial parietal, L. PFC=lateral prefrontal, SM=sensory/motor, B=bilateral, L=left, R=right, D=depth electrodes, E=ECoG) (d) Example depth electrode for each of the 4 patients and ECoG localizations for 2 patients.

Intracranial recordings in epilepsy patients provide both temporal and spatial high-resolution multilevel neural measurements of single-neuron activity and intracranial EEG (iEEG, Figure 77). In order to additionally provide large-scale network view, we obtained whole brain fMRI activation in the same patients, as well as in 55 healthy volunteers. We used two interactive computer games: first, the Punishment, Reward & Incentive Motivation (PRIMo) game (Gonen et al., 2016) (described in Figure 78), and the Risky Choice Domino (RCD) game (Gonen et al., 2012). Patients played both games during their implantation period allowing the collection of neural activity during motivational processes. To our knowledge, this is the first demonstration of a comprehensive dissection of the motivational process into its different stages in such high-resolution. Providing such

insights may guide interventions corresponding to a pathological condition involving a specific stage.

Description of recordings and data

Invasive recordings were obtained from 10 neurosurgical patients with pharmacoresistant epilepsy who are implanted with intracranial depth electrodes (7 patients) or subdural electrodes (ECoG, 3 patients) for 7-10 days, to determine the seizure-onset zone for possible surgical resection (Figure 77). The Benkhe-Fried depth electrodes used include two types of contacts: Each electrode contains eight 1.5-mm-wide macro-contacts along the shaft allowing the recording of electrophysiological currents representing a summation of local currents (Buzsáki et al., 2012b). In addition, eight 40 µm micro-wires situated at the tip of each electrode will record extracellular EEG and single-unit activity (Fried et al., 1999) (Figure 77). Thus single cell and LFPs and iEEG was collected from mesial structures highly relevant for mediating motivation related processed including the hippocampus, amygdala, cingulated cortex and medial prefrontal cortex (mPFC). The iEEG recordings were collected from these mesial regions but also covering lateral cortical sites primarily in the temporal and frontal lobes. ECoG recordings (3 patients) were collected using 2 mm diameter macro contacts with 8 mm spacing between adjacent electrodes allowing mapping of lateral brain regions. Figure 77 describes the electrodes used and the brain regions recorded for each patient.

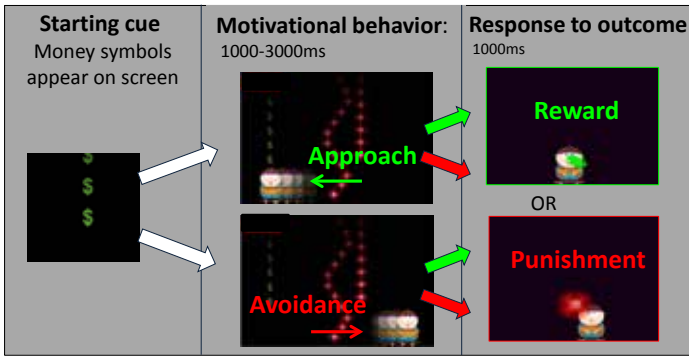


Figure 78: The Punishment, Reward & Incentive Motivation (PRIMo) game

The goal of the game is to earn money by catching coins and avoiding balls. There are two ways to gain or lose money: controlled – the player actively approaches coins and avoids balls; and uncontrolled –random coins and balls colored differently hit the player. Adopted from Gonen et al., 2016.

The database includes the following: (a) A summary table of electrode location in xyz Montreal Neurological Institute (MNI) coordinates and their atlas labels using the neuromorphics atlas (http://www.oasis-brains.org/, http://Neuromorphometrics.com/ segmenting the human brain to 207 brain regions). This includes microwires locations which can detect single cell recordings and macro electrode locations recording local field potentials (LFP). (b) A table describing neuronal firing showing a significant increase or decrease in firing rate in response or expectancy for a paradigm induced event (for example: an account of neurons in different brain location responding to control punishment or reward in the PRIMo paradigm, or responding to anticipation in the RDC game). (c) Matlab matrices describing the power of LFP signal at the different frequency bands at the different spatial locations responding to paradigm induced events.

A **Dataset Information Card** has been completed (see DIC Task T3.2.4 "Intracranial recordings in motivational paradigm").

Location of our data storage: The data will be hosted by the Center for Brain Functions at Tel Aviv Medical Center (TASCM) at url: http://fmri-tlv.org/tomer.html.

Provenance of the data: All data were collected at TASMC in a collaborative effort between the Center for Brain Functions and the Functional Neurosurgery unit. The database will include a table describing and tracking data provenance.

Self-analysis of the value and completeness of our data: This database will be regarded as complete when including final localization of all channels and neural responses at the different scales and paradigm conditions from 12 patients. We are now collecting the final two patients.

Indication of who has used this data so far and for what: The data have only been used by TASMC. A manuscript detailing micro and meso-scale neural responses to motivational states and values is in preparation. A manuscript detailing macro scale fMRI responses to motivational decision making (in the PRIMo game) in a healthy population has recently been accepted ((Gonen et al., 2016), HBP acknowledgement).

Main Results

The purpose of the PRIMo task was to illuminate the neural dynamics underlying motivational decision making by deconstructing it into stages; incentive cue, decision under conflict (to approach or avoid the cue), action behavior (approaching or avoiding) and response to outcome (reward or punishment) (see Figure 78 for trial structure). A cascade of neural activations was observed using intracranial recordings allowing a conscientious account of the neural dynamics involved on motivational decision making. The following describes the neural activity detected at different relevant brain regions at five stages of motivational decision making. We focused on 5 brain regions which are known to play significant roles in this process: the amygdala, hippocampus, parahipocampal area, ventral medial prefrontal cortex (vmPFC) and dorso-medial prefrontal cortex (dmPFC). Here we present finding from the PRIMo paradigm which allow the full tracking of motivational decision making stages in an ecological scenario. Findings from the RCD paradigm are presented as a complementation for the anticipation period. Figure 77 details the amount of neurons and iEEG channels used for this analysis.

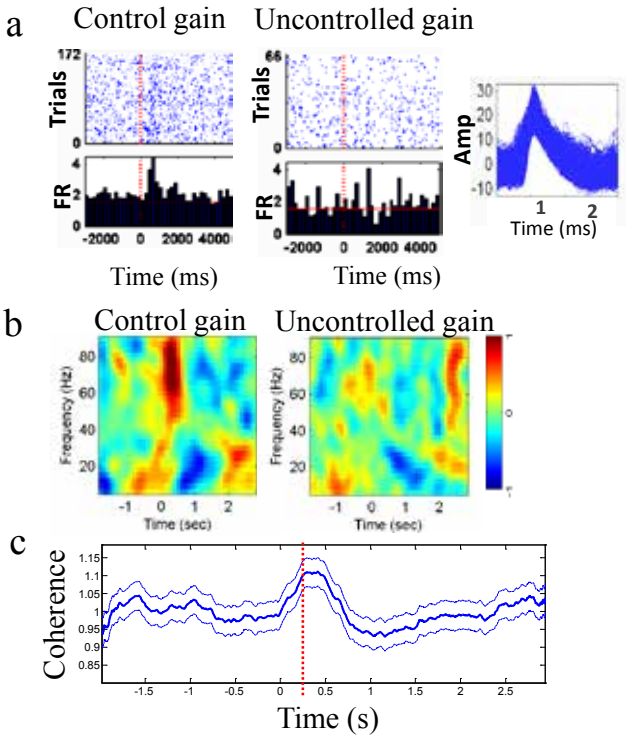


Figure 79: Neural response to gain cue appearance

(a) An example of neuron in the parahippocampus presenting an increase in FR following cue appearance of control reward trials. (b) Mean regional spectrograms. Time-frequency analysis of iEEG revealed an increase in gamma power 100 to 500ms post cue appearance [p<0.001 bootstraping, N=1000] (example shown from the vmPFC). This effect was stronger for incentive (control) trials. (c) Gamma (40-80Hz) phase coherence increase following gain control appearance between medial frontal and mesial temporal regions [t(32)=4.97,p<0.001]. This effect was not found for lower frequencies.

The PRIMo: Response to appearance of incentive cue [0-500msec following appearance]. A positive ($) or negative (ball) cue appeared at the top of the screen. A strong amplification of gamma band following the appearance of the cue signaling its detection and processing (Figure 79). While all mesial brain regions evaluated presented this increase, it was most dominant in the Amygdala. This amplification was stronger for incentive trials as compared to non-incentive trials (i.e. non-control trials). This gamma band amplification was accompanied by increased phase coherence (synchronization) between mesial temporal and medial frontal regions and increased or decreased firing rates of neurons in all regions. Neurons in all 5 regions showed more sensitivity to the punishing cue (24% mean responsivity – as measured by the percent of neurons per region, from the total neurons recorded, with a significant alteration in FR) compared to the appearance of the rewarding cue (9% mean responsivity). This suggests that incentive cues are more salient and demand more fronto-limbic resources than the non-incentive cues; with incentive punishing cues demanding the most resources than incentive rewarding or non-incentive cues.

Decision making under conflict [300-800msec after cure appearance]. The positive cue has to be achieved but negative cues may stand in the way inducing goal conflict between approach and avoidance behavior. A strong increase in delta and theta bands (2 to 7Hz) after positive cue appearance in iEEG channels within the hippocampus (Figure 80). This increase in stronger in conditions with high goal conflict (2 or more bombs in the way)

compared to low goal conflict (0 or 1 ball).This is in compatibility to the a neurobehavioral model of motivation known as the Reinforcement Sensitivity Theory (RST), suggesting that increase in hippocampal theta in gating the recursive hippocampus-vmPFC signaling, in order to resolve high goal conflicts by comparing between the current state, previous knowledge and expected consequences, for the sake of adaptive behavioral selection.

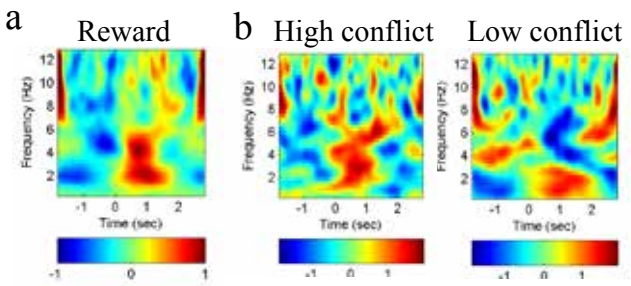


Figure 80: iEEG time frequency analysis for decision and action

(a) Average of low frequency (2-6Hz) power increase in the Hippocampus in response to control appear gain at 300-1000ms (N=11 iEEG channels). This effect was not found for other brain regions (b) Classifying trials to high and low goal conflict revealed that this effect is stronger for high conflict (left) compared to low conflict (right). An example hippocampal electrode from D3 is shown.

Action behavior [600-1200ms after cue appearance]. Approach or avoid behavior takes place. A decrease in gamma power in mPFC and amygdala but not in the hippocampus appears during this time (Figure 81). Similar pattern can be seen in the 46 healthy subjects' results showing BOLD reduction in the amygdala and vmPFC during approach behavior (vs. anticipation); while increase in BOLD signal is clearly evident in action related regions such as the brainstem, dorsal striatum, and premotor and motor cortices (Figure 81). Taken together it seems like the limbic-PFC regions are reducing activity in order to allow for action preparation and execution to take place. Similar finding though less robust were evident in patients.

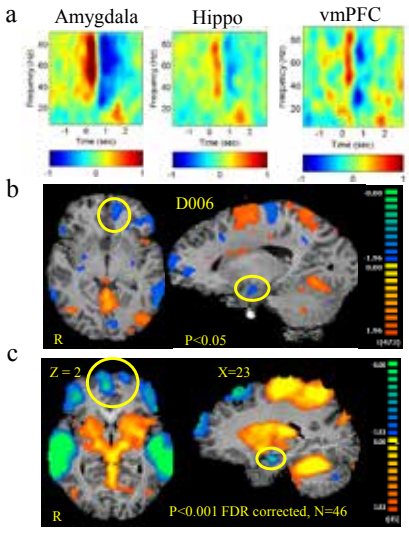


Figure 81: Multi scale response to decision and action

(a) Average high frequency spectrogram of iEEG channels: 8 Amygdala, 11 hippocampus and 4 vmPFC channels. A decrease in gamma band appears around 600ms following cue appearance. [p<0.001 bootstraping, N=1000 for amygdala and vmPFC]. (b-c) fMRI activations during behavior. b. The contrast of control vs. non-controlled: Reduction in activity in the vmPFC as well as the amygdala is demonstrated in patient D006 (p=0.05 uncorrected). c. fMRI from 46 healthy subjects shows a similar pattern of BOLD reduction in the amygdala and

vmPFC during approach behavior (vs. uncontrol); while increase in BOLD signal is clearly evident in action related regions such as the brainstem, dorsal striatum, and premotor and motor cortices.

Response to reinforcing outcome [0-500ms following outcome]. Positive or negative cue hits the player, resulting with reward (additional 5 points to player's score in the game) or punishment (5 points reduced from player's score in the game). Increase in gamma band power in the amygdala and mPFC but not in the hippocampus appeared following outcomes. (Figure 82). This amplification was stronger for incentive trials (controlled) as compared to non-incentive trials (non-control). This gamma band amplification was differently distributed in response to reward and punishment. The amygdala and dmPFC showed a higher increase following punishment outcome (compared to reward) while the hippocampus and vmPFC showed a higher increase following reward outcome (compared to punishment). This response was accompanied altered FR of neurons in the different regions. Concordantly, in response to punishment outcome, the amygdala and dmPFC showed higher neural reactivity (55%) compared to the hippocampus and vmPFC (39%), and the opposite trend was observed in response to reward outcome (16% reactivity in the amygdale and dmPFC and 32% reactivity for the hippocam pus and vmPFC). Interestingly, consistent with previous studies (Morrison and Salzman, 2010), neurons in the Amygdala tend to decrease FR in response to punishment, while an increase in FR was more evident in other regions (Figure 82). Additionally, increased phase coherence (synchronization) between mesial temporal and medial frontal regions was observed during this time window (Figure 82). Data from 55 healthy subjects revealed compatible large-scale network segregation, with dorsal mPFC, SMA and motor cortices; along with brainstem and thalamus showing greater sensitivity to controlled punishment; while ventral mPFC, including the vmPFC and sgACC; as well as the hippocampus and para-hippocampal gyrus showing greater sensitivity to non-controlled-rewarding trials. Similarly to the appearance phase, this suggests that the saliency of incentive, punishing outcomes is greater than incentive reward and from non-incentive outcomes. These results were also evident in patient's fMRI data, showing a pre-frontal segregation, where dmPFC showed greater activation during incentive-punishment trials, while the vmPFC showed greater activation during non-incentive rewarding trials (Figure 82).

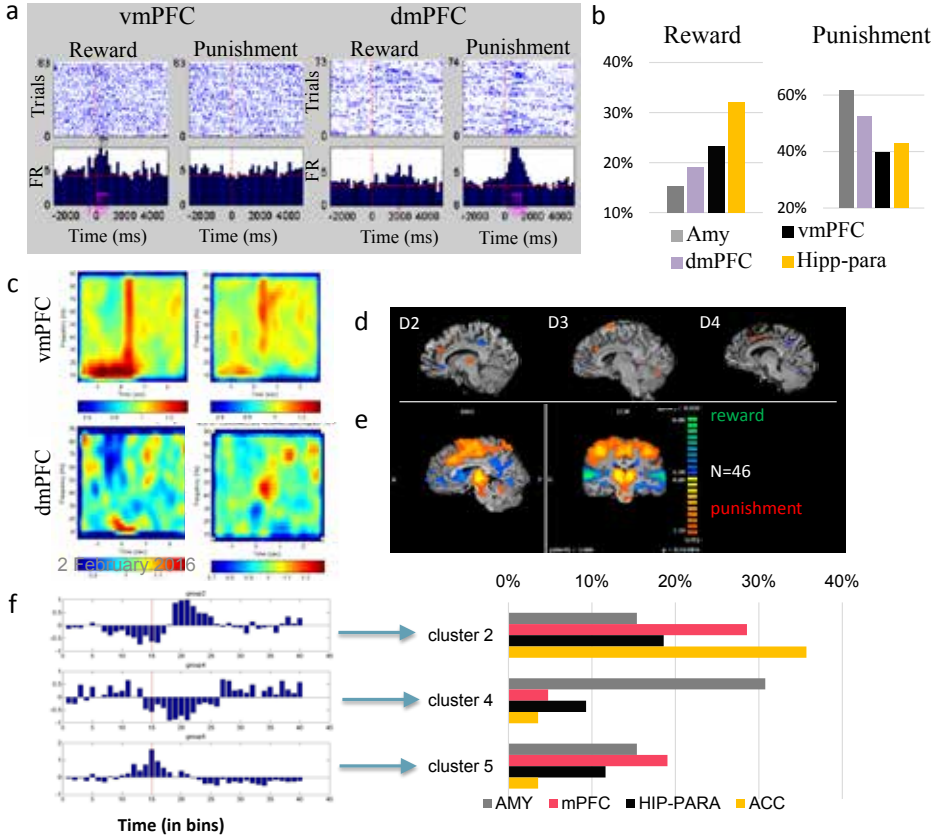


Figure 82: Multi-level neural response to outcome

(a) An example of two neurons in the vmPFC (left) and dmPFC (right) which increase their FR following the outcome (reward or punishment respectively). Top- raster plot for the reward and punishment conditions. Bottom – Peri-Stimulus Histrograms. (b) Time-frequency analysis (averaged over all amygdala and hippocampus iEEG channels). An increase in gamma band power is observed in the amygdala peaking at 300ms following outcome but not in the hippocampus. (c) Summary of neuron reactivity to outcome shows, in response to punishment outcome, AMY and dorso-mPFC showed higher neural activity compared to the hippocampus and ventral-mPFC and the opposite trend was observed in response to reward outcome. (d) Response to outcome – fMRI in patients. Conjunction analysis of: controlled vs. non-controlled & punishment vs. reward revealed similar segregation of pre-frontal regions to with dmPFC/ dACC and SMA more sensitive to controlled and punishment trials, whereas vmPFC was more sensitive to non-controlled and rewarding trials (demonstrated on patients D2 and D4; p=0.05, uncorrected). (e) fMRI contrast maps punishment vs. reward from 50 healthy subjects show a similar pattern in the brain regions discussed. (f) Neurons showing a significant alteration in FR in response to punishment were clustered according to their PSTHs. Notice that the cluster showing a decrease in FR is dominated by amygdala neurons.

The RCD paradigm: Anticipation. We focused on anticipation period in the RDC game to complement the PRIMo limitation during this anticipation to outcome stage. In this gambling game a player makes a decision between low risk (matching chip) and high risk (non-matching chip). Neurons in the mPFC were found to alter their FR within anticipation (Figure 83). In iEEG channels we found a wide band power increase within this period in the amygdala and mPFC. The Amygdala showed a higher gamma increase for the risky condition, possibly reflecting the anticipation of punishment. This corresponds to our original fMRI finding (Kahn et al., 2002) showing increased amygdala activation during anticipation periods that follow risky choices (thus high probability of punishment).

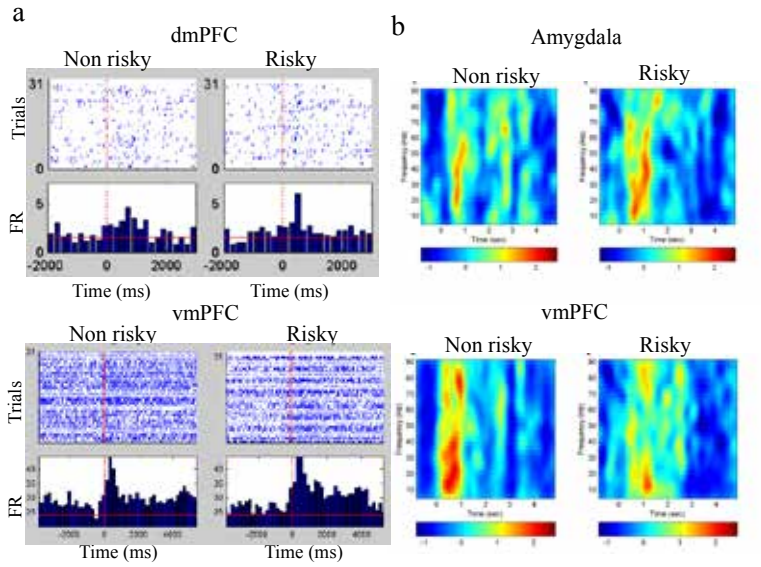


Figure 83: Anticipation to outcome in the RCD game

(a) An example of two neurons (D5) in the dmPFC (top) and vmPFC (bottom) show increased FR during the anticipation period following risk or no-risk choices. (b) Example spectrograms (from patient D7). In the Amygdala, a wide band increase in power is stronger following risky decision (anticipation to punishment) vs. non risky decision (anticipation to reward). The opposing trend is shown in an example vmPFC channel.

Summary

Taken together, our results reveal a cascade of neural activations underlying the dynamics of motivational decision making (see a hypothetical model that is inspired by our fMRI and intracranial recording findings and from previous work in Figure 84). Immediately following the appearance of an incentive cue, increased activity was evident in limbic (amygdala, hippocampus) and prefrontal nodes, along with increased fronto-temporal synchronization. Such integrated effort is expected during detection and evaluation of a significant stimulus such as reinforcing cue which carries motivational saliency (Bressler and Menon, 2010). Subsequently, a behavioral decision to approach or avoid is required, previously indicated to be gated by increased activity in hippocampal theta band (Gray and McNaughton, 2003). This was demonstrated in our data, mostly under high goal conflict, known to increase the complexity of the motivational decision. Following decision, approach behavior was characterized by decreased activity in the limbic system and in PFC, though most robust in the amygdala. This reduction in activity probably marked shift in resources towards action facilitation systems such as the striatum and motor related regions (Pessiglione et al., 2007). Similar results were shown in patients' fMRI data (Figure 81), as well as in a separate fMRI study of healthy subjects (Gonen et al., 2016), revealing wide-spread saliency and motor preparation activation following incentive cue appearance (combining appear, decision and action due to poor temporal resolution of the fMRI). Finally, in response to incentive outcome, amygdala and dmPFC showed increased activity and were more sensitive to punishment, while the hippocampus and vmPFC showed greater sensitivity to reward. This too was compatible to our healthy-subjects fMRI data, showing more distributed response, yet the hippocampus-vmPFC vs. amygdala-dmPFC segregation in response to reward vs. punishment outcomes (respectively) was clearly evident (Figure 82). This is the first multi-scale neurobehavioral portraying of motivational decision making in humans. Our findings may guide neural based interventions corresponding to pathological conditions involving specific motivational stages.

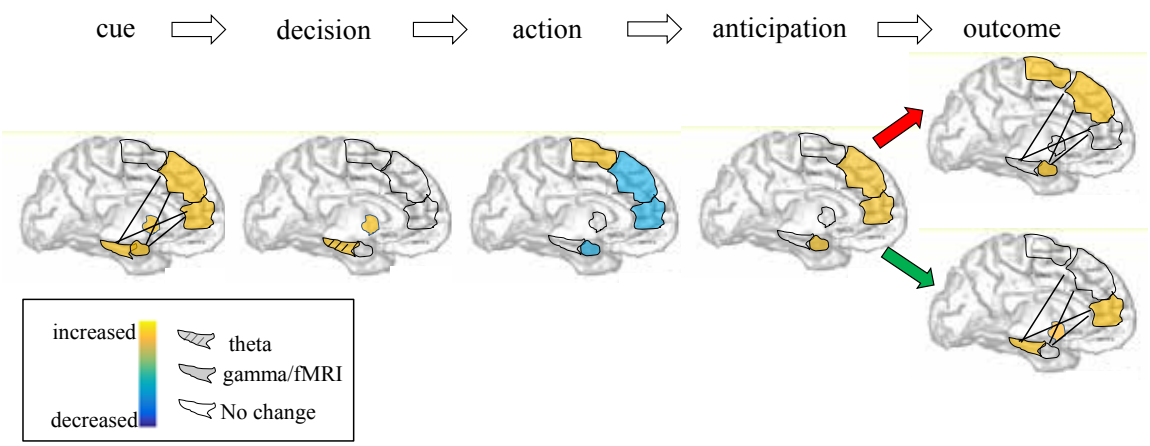


Figure 84: Scheme of motivational decision making model

Learning and Memory

Coordinated by Yadin Dudai

Our goal in WP3.3, *Learning and Memory*, was to identify and map the neuronal circuits and hence elucidate the cognitive architectures (as defined by Dehaene, Dudai and Konon, *Neuron* 2015) of selected human learning and memory systems. This was achieved in 3 tasks that were deemed by us of particular relevance to large scale human brain simulations and also to future neurocognitive robots the design and construction of which may benefit from such simulation: T3.3.1, Skills and Habits (UHAIFA); T3.3.2, Memory for Facts and Events (WIS, EKUT); and T.3.3.2, Working Memory (UMU, WIS). The milestones of our research followed those defined in the DoW, namely, developing and validating protocols for localizing the aforesaid cognitive architecture using fMRI; generating the datasets; and subsequently model the cognitive architecture of the distinct cognitive function. As expected, iterations were made in each task in validating the protocols, reacquiring data on the bases of the initial experiments, and modifying and improving the models.

T3.3.1 focused on overnight procedural memory consolidation, based on a motor repetition protocol, and generated a model assigning roles for overnight modifications in intrinsic motor cortex (M1) connectivity as well as extrinsic connectivity to the basal ganglia. T.3.3.2 focused on three sub-tasks. First, the cognitive architecture that subserves the encoding and initiation of episodic (event) memory. This task discovered the role of role of distinct brain circuits in peri-encoding (i.e., a few seconds before and a few seconds after encoding). Furthermore, the engagement of these circuits predicts subsequent memory in a realistic episodic memory protocol. This approach was able to further tease apart a fast shift between encoding- and retrieval mode of the hippocampus - a finding of special importance to modelling hippocampal role in episodic encoding and retrieval. In the second sub-task, the role of sleep in episodic consolidation has been investigated and a detailed dynamic mass model implicating distinct sleep phases and neuronal states has been generated. In the third sub-task, the role of emotion in encoding events and facts was analyzed. Here, using a new behavioral paradigm, it was shown that anxiety leads to wider generalization for loss- as well as gain-conditioned stimuli, and the circuits underlying this behavioral modification were identified, including amygdala, dorsal anterior cingulate cortex (dACC) and the Putamen. T.3.3.3 focused on short-term maintenance of conscious and non-conscious information. Using delayed matching-to-sample, evidence was found for sustained mnemonic effects indicating non-conscious working memory. Two points are of note: first, classically, working memory was considered as a cognitive task under executive, attentional control, and the new data reinforce the non-conscious, 'automatic' contribution' of particular interest to brain simulations that do not necessarily assume emergence of consciousness; second, the data are consistent with the model generated in T3.3.2 above, in which immediate post-acquisition binding in a non-conscious working memory buffer is critical for subsequent episodic memory.

All in all, WP3.3 hence generated multiple cognitive architecture models at various levels of realization in the human brain (from cellular-circuits to systems). These models can serve as boundary conditions as well as guides for simulations of key mnemonic functions in the human brain.

3.1 The consolidation and Transformation of memory

Review of the cognitive architecture for the consolidation and transformation of memory

Yadin Dudai, Avi Karni, Jan Born "The Consolidation and Transformation of Memory", Neuron, Volume 88, Issue 1, p20–32, 7 October 2015

Abstract

Memory consolidation refers to the transformation over time of experience-dependent internal representations and their neurobiological underpinnings. The process is assumed to be embodied in synaptic and cellular modifications at brain circuits in which the memory is initially encoded and to proceed by recurrent reactivations, both during wakefulness and during sleep, culminating in the distribution of information to additional locales and integration of new information into existing knowledge. We present snapshots of our current knowledge and gaps in knowledge concerning the progress of consolidation over time and the cognitive architecture that supports it and shapes our long-term memories.

Data set 1: Short-term cortical modulation by task repetition as signatures of procedural memory consolidation

Task T3.3.1 - Avi Karni (UHAIFA)

Overview of empirical work

The strategic question addressed: whether and how are motor skill consolidation processes reflected in the modulation of neuronal responses to task repetition.

The aim of the studies undertaken in Task 3.3.1 (skills and habits) was to test cortical responses to repeated action as brain signatures of accumulating experience, plasticity and procedural memory consolidation, in motor skill learning in young adults. To this end, data from an fMRI experiment addressing cortical dynamics in movement repetition for two movement sequences composed of identical component finger movements (a trained sequence – over which subjects have slept, and a novel sequence). The results show that short term but robust modulations of the primary motor cortex activity, its intrinsic connectivity as well as the M1's extrinsic connectivity to the basal ganglia, reliably reflect the individual's level of experience with a sequence of movements. We propose that M1 not only generates movements but also serves as a hub for a motor working memory system: wherein transient stabilization of activity upon sequence repetition reflects short-term familiarity with a novel sequence of movements. A temporarily stabilized network in cortex and striatum may promote an integrated representation of the new movement sequence (i.e., the movement syntax). Importantly, when a well-consolidated movement sequence is repeated, the M1 - striatum functional connectivity decreases upon repeated performance, as one would expect in an "automatic" response. Averaging over single events or blocks may not capture the dynamics of motor representations that occur over multiple time-scales; transient but consistent changes in motor cortex activity and connectivity to repeated experience are key elements in the dynamic representation of actions in cortex. Two papers (Gabitov et al., 2014, 2015) present this work. An additional aspect that was addressed was an analysis of the neural architecture supporting the ability to transfer (generalize) practice-related performance gains (skill, procedural knowledge) from the trained limb to the other. The data suggest that a critical hub is the pre-motor cortex in the trained hemisphere; however, mnemonic knowledge transfer is driven by the trained M1's modulations. This work has recently been published (Gabitov et al., 2016)

An additional behavioral and fMRI brain imaging study, addresses motor cortex plasticity driven by visual input (action observation). The data suggest that while repeated can be highly effective in the acquisition of skilled motor performance, as well as in initiating memory consolidation processes, skill memory from action and memory from observation may not overlap in the brain representations of learned movement sequences. A paper presenting the behavioral results has been accepted for publication (Maaravi-Hesseg, Gal & Karni, Learn & Mem., *accepted for publication*). The imaging data suggest that M1 was only activated in the actual performance of both movement sequences but not during their viewing and we argue that the lack of M1 activation in the observation condition may explain the lack of behavioral interaction between the two modes of training.

General design and methods

These are described in details in three published papers (Gabitov et al, 2014, 2015). In brief the approach was to train young healthy adults in the performance of an explicitly instructed movement sequence (T-FOS, A or B) and retest (and comparing to an untrained sequence, U-FOS) at 24 hours post-training for overnight, consolidation phase gains (Figure 85). Whole brain BOLD fMRI data at 3T was acquired in the overnight session while participants were tapping the sequences (T-FOS, U-FOS) repeatedly at a paced rate and

the data were analyzed for modulations of the evoked BOLD signals as signatures of overnight procedural memory consolidation, as well as for changes in functional connectivity dynamics as a function of mnemonic processes.

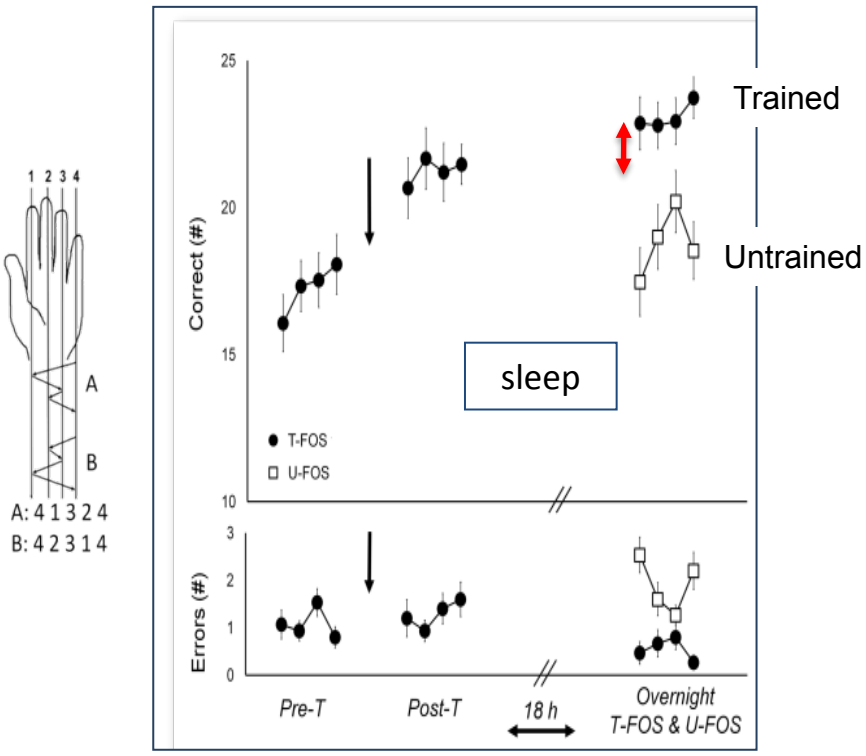


Figure 85: Graphical depiction of the two movement sequences (left) and the main behavioral outcomes (right, speed - upper panel, accuracy - lower panel).

The results indicate the expression of overnight "offline" performance gains in the performance of the T-FOS (red double arrow) as well as a clear advantage in the performance of the T-FOS over the U-FOS. (Gabitov et al., 2014).

Results

- Behaviourally, practice lead to robust "offline" gains in speed and accuracy for the trained finger movement sequence; these gains were expressed after overnight consolidation interval (Figure 85).
- The imaging data showed brief but robust modulations (repetition enhancement, RE) of M1 activity correlated with the behavioral signature of consolidation (the overnight, "offline" movement speed gains) (Figure 86) as well as that M1's intrinsic connectivity and M1's extrinsic connectivity to the basal ganglia (Figure 87), reliably reflect the individual's level of experience with a sequence of movements.

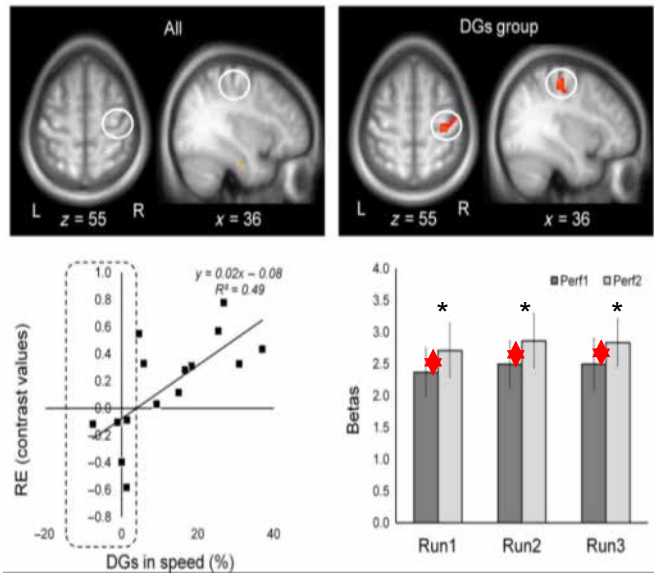


Figure 86: The imaging findings. Robust repetition enhancement (RE) effects occurred for repeating well-trained movement sequences. RE was correlated with the behavioral signature of consolidation - overnight, "offline" movement speed gains. (Gabitov et al., 2014)

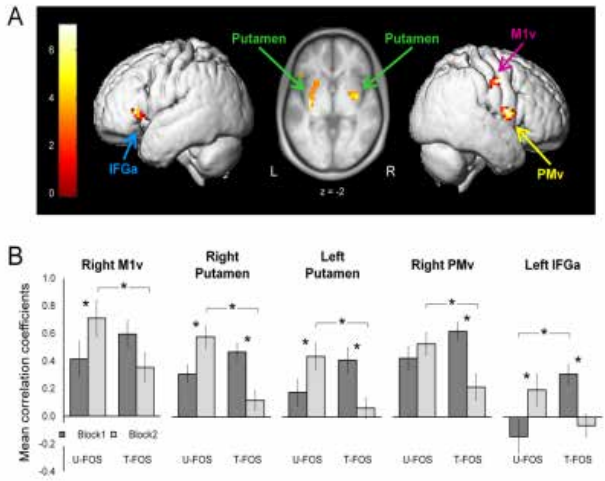


Figure 87: The imaging findings. Connectivity between M1 and basal ganglia increased for repeated new sequences; but decreases for consolidated sequence (Gabitov et al., 2015a).

A subsequent study addressed the neural architecture supporting the ability to transfer the skill to an untrained limb (Gabitov et al, 2015). Participants were scanned during the performance of the movement sequence, intensively trained a day earlier and also in the performance of a similarly constructed but previously untrained sequence (Figure 85, T-FOS and U-FOS, respectively) with both sequences performed with the untrained hand.

Results:

- The performance of the T-FOS was accompanied by larger activity in both M1s. The differential responses in the 'trained' M1, ipsi-lateral to the trained hand, were correlated with the experience related differences in the functional connectivity

between the 'trained' M1 and 1) its homologue and 20 the dorsal pre-motor (PMd) cortex within the contra-lateral hemisphere.

- No significant correlations were evident between experience related differences in M1 – M1 - PMd connectivity measures.

These results suggest that the transfer of sequence specific information between the 2 primary motor cortices is predominantly mediated by excitatory mechanisms driven by the 'trained' M1, via two independent neural pathways (Gabiotov et al., 2015).

An additional behavioral and fMRI brain imaging study (N=20) addressed the acquisition of skill, in the FOS task, from actual practice (as administered in the above studies, see training protocol in Gabitov et al., 2014 vs. learning from observation, i.e., in a condition wherein participants were asked to follow by observing only the video-taped repetition of the FOS (Figure 85). To this end we ran two behavioral experiments and a brain imaging study of motor mnemonic representations driven by visual input (action observation) compared to actual, physical, performance.

PMd Rt Act [30 -10 55], r = 6mm
PMd Lt Act [-21 -4 52], r= 6mm

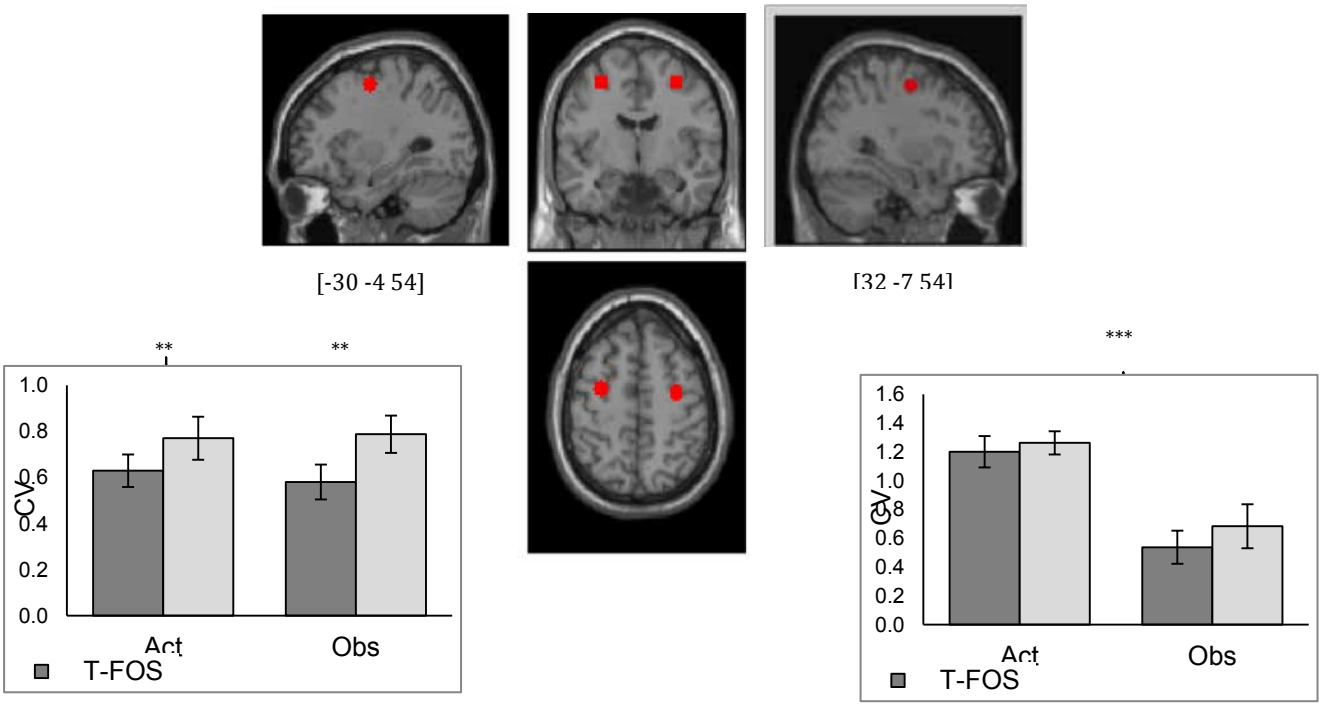


Figure 88: Differential activity induced by actual performance (Act) and observation (Obs) in the pre-motor cortices of the 2 hemispheres (respectively) in the corresponding dorsal pre-motor cortices (PMd), for the T-FOS after an overnight consolidation phase and the newly introduced U-FOS.

Importantly, the PMd bilaterally did not distinguish between the actually performed sequences and observed sequences (i.e., no differential activity for action vs. observation). Unlike the expectation from some studies of the human 'mirror neuron' system, the ventral PM did not show differential responses for the two sequences (not shown).

Results:

- The behavioral data suggest that: (1) both training by executing and by observing movements can improve task performance as well as trigger skill consolidation processes as reflected in the expression of robust delayed "offline" gains in the performance of the trained movement sequence. However, (2) consolidation could be blocked by ensuing action but not by observation, indicating that skills acquired in doing or observing do not necessarily overlap in terms of their brain representation. A paper is in review (Maaravi-Hesseg, Gal & Karni, *in review*).
- M1 was only activated in the actual performance of both movement sequences but not during their viewing.
- Activity in the PMd, bilaterally, did not distinguish between the actually performed sequences and observed sequences (i.e., no differential activity for action vs. observation) (Figure 88). However, no such differential activity was found in the PMv (implicated in studies of the putative human 'mirror neuron' system.
- The SMA showed sequence specific activity differences only when the movements were observed but not when the same movement sequences were performed.

We therefore propose that although repeated observation can lead to both performance gains and a mnemonic representation, skill memory from action and memory from observation may not overlap in terms of brain representations.

A perspective and proposal of an integrative reframing of the notion of memory consolidation as a generative process has been advanced by members of SP3.1 (see Dudai, Karni & Born, 2015).

Conclusion

We propose that the modulation of activity in a given brain area (M1) by task repetition reflects learning and overnight procedural memory consolidation. Specifically, brief but robust modulations of M1 activity as well as M1's intrinsic connectivity and M1's extrinsic connectivity to the basal ganglia, reliably reflect the individual's level of experience with a sequence of movements. M1 serves as a hub for a motor working memory system: wherein transient stabilization of activity upon sequence repetition reflects short-term familiarity with novel movement sequences (i.e., the movement syntax). However, this hub is under-engaged in practice through observation and thus, although training by observation can trigger skill consolidation processes, the knowledge attained from actual practice as compared to that attained from the observation of the same actions may be qualitatively different.

These results are therefore of high relevance to the modeling of learning (experience) related changes and mnemonic processes in brain areas, presumably at columnar level, as they suggest a specific, and dynamic set of signatures for repeated experience vs. novelty. The results are also of importance in the improvement of intervention programs for the acquisition of skills.

The following papers were included in the HBP support period:

1. Gabitov, E., Manor, D., & Karni, A. (2014). Done that: Short- term Repetition Related Modulation of Motor Cortex Activity as a Stable Signature for Overnight Motor Memory Consolidation. Journal of Cognitive Neuroscience, 26(12), 2716-2734. http://doi.org/10.1162/jocn

2. Gabitov, E., Manor, D., & Karni, A. (2015). Patterns of Modulation in the Activity and Connectivity of Motor Cortex during the Repeated Generation of Movement Sequence. Journal of Cognitive Neuroscience, 27(4), 736–751. http://doi.org/10.1162/jocn
3. Dudai, Y., Karni, A., & Born, J. (2015). The Consolidation and Transformation of Memory. Neuron, 88(1), 20–32. http://doi.org/10.1016/j.neuron.2015.09.004
4. Gabitov, E., Manor, D., & Karni, A. (2016). Learning from the Other Limb's Experience: Sharing the "Trained" M1's Representation of the Motor Sequence Knowledge. Journal of Physiology, 1, 1–39. http://doi.org/10.1113/JP270184.Thi
5. Maaravi-Hesseg,R., Gal,C., & Karni, A. (2016) Not quite there: skill consolidation in training by doing or observing. *Learn.Mem* (*In the press*).

Talks

Data and models involving our WP were delivered by Avi Karni at the Symposium on Motor Skills, the International Meeting on Cognitive & Neuro-cognitive aspects of Learning Abilities & Disabilities, (Haifa, May, 2015); The Rehabilitation Science and Technology Update 2016 (Rishon Letsion, February, 2016); The 8th Haifa Forum for Brain and Behavior (Haifa, February, 2016) and the Neuroscience Seminars, University of Maastricht (Maastrict, February, 2016).

A **Dataset Information Card** has been completed (see DIC Task T3.3.1 "Learning and memory: motor skill consolidation and intermanual transfer").

Data Provenance: The behavioral and fMRI data were acquired and analyzed by Ella Gabitov, David Manor, Rinatia Maaravi Hesseg & Carmit Gal at the University of Haifa, the fMRI unit, C. Sheba Medical Center, Tel-Hashomer and fMRI unit, Rambam Medical Center.

Data Location: The data were deposited on a server at:
https://openfmri.org/dataset/ds000170/

Data set 2: Cognitive architecture of the initiation of systems consolidation

Task T3.3.2 - Yadin Dudai (WIS)

Our goal in this task was to contribute to the understanding of the cognitive architecture of the initiation of memory consolidation in the human brain. An abundance of research in the field of human memory has focused on the encoding and subsequent consolidation of episodic memory. Yet very little is known about the transition between the two, encompassing the first seconds following the inception of the memory trace. In a recent set of studies, we developed a new paradigm to specifically target memory processes elicited at the offset of episodes (Ben-Yakov and Dudai, 2011). By presenting participants with short movie clips of varying lengths (4-16s), intercalated with brief rest periods, we identified brain regions displaying memory-predictive activity at event offset. These consisted of regions in the striatum, hippocampus and cerebellum (all bilaterally), all demonstrating a response that was time-locked to the clip offset and predictive of subsequent memory for the clip's gist (Figure 89).

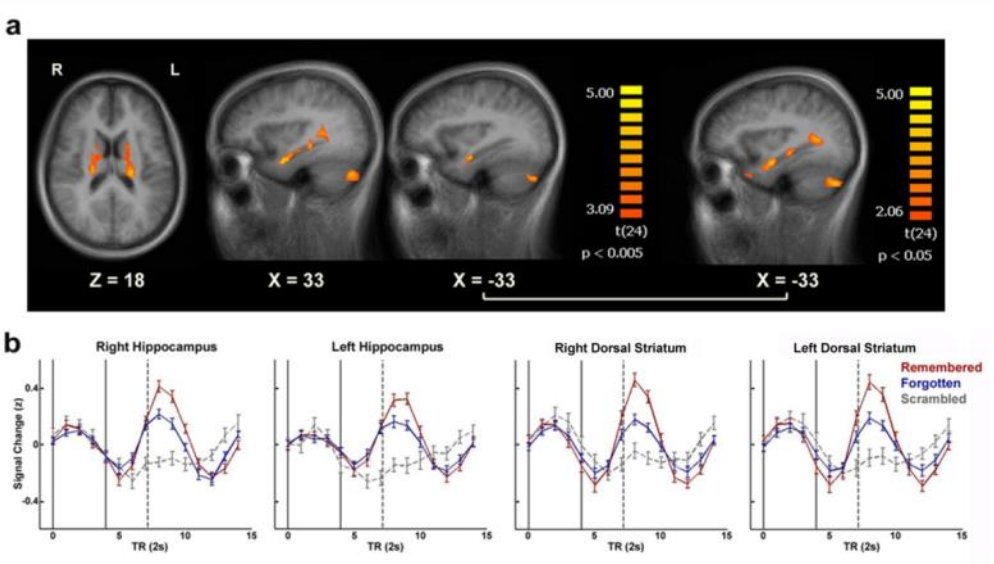


Figure 89: Regions demonstrating memory-predictive activity at event offset

(a) Regions showing a significant difference in BOLD activity at the offset of subsequently remembered vs. subsequently forgotten clips in conjunction with a positive response at the offset of forgotten clips relative to baseline ($p<0.005$ for each contrast, uncorrected, minimal cluster size 5 contiguous functional voxels, GLM with a random effects group analysis, n=25). Data are shown on axial and sagittal slices of the group-average brain. On the right, a slice including the left hippocampus is shown with the same contrast at a more relaxed threshold ($p<0.05$). (b) Mean group BOLD signal (after z-scoring each time-course) during and following remembered, forgotten and visually scrambled clips. Error bars show standard error of the mean. The black lines indicate the onset (left line) and offset (right line) of clip presentation, while the dashed line indicates the mean onset of the following clip. Results are shown for the bilateral hippocampus bodies and bilateral dorsal striatum (dorsal caudate nucleus). Adapted from Ben-Yakov and Dudai 2011.

Based on the initial observation, we hypothesized that the hippocampal response is triggered by the closure of the episode, as indicated by a salient change in the stream of information. To test this, we presented participants with pairs of movie clips in immediate succession (Ben-Yakov et al., 2013), and found that the hippocampus responded at the offset of each clip (Figure 90). This design also enabled us to study the brain substrates of retroactive interference, a behavioural phenomenon in which presentation of two stimuli in succession impairs memory for the first of the two. We found that attenuation of the posterior hippocampal response at the offset of the first clip corresponded with the behavioural effect (Figure 91), suggesting that disruption of post-event processing in the posterior hippocampus may account for this phenomenon.

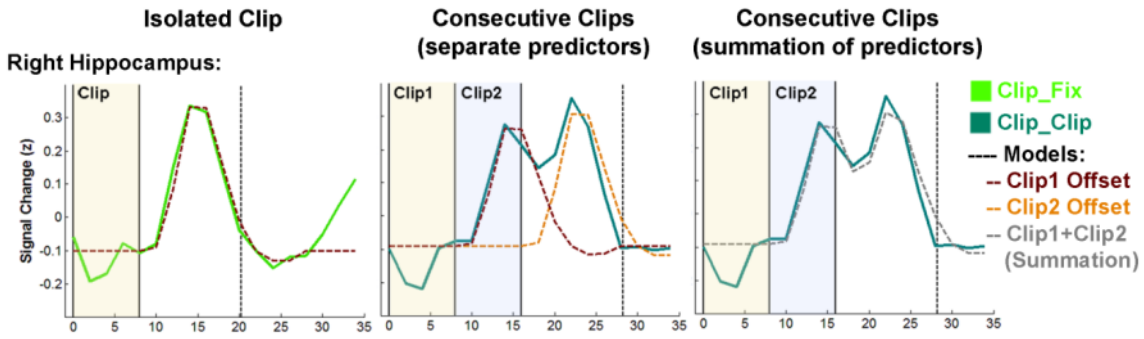


Figure 90: Double-peaked hippocampal response to consecutive clips

Left panel: the average response (average of the z-scored data over trials and participants) to presentation of a single clip (Clip_Fix), plotted with the fitted model of response at clip offset (dashed line). Middle panel: the average response to two consecutive clips (Clip_Clip) plotted with the fitted model predictors of a response at the first clip offset (brown dashed line) and at the second clip offset (dashed orange line). Right panel: same as the middle column, with the summation of the separate model predictors (gray dashed line), demonstrating the fit of the data to a double-peaked model. In all plots, the solid vertical lines represent the onsets and offsets of the clip(s) while the dashed vertical line represents the average end of the fixation screen. Adapted from Ben-Yakov, Eshel and Dudai 2013.

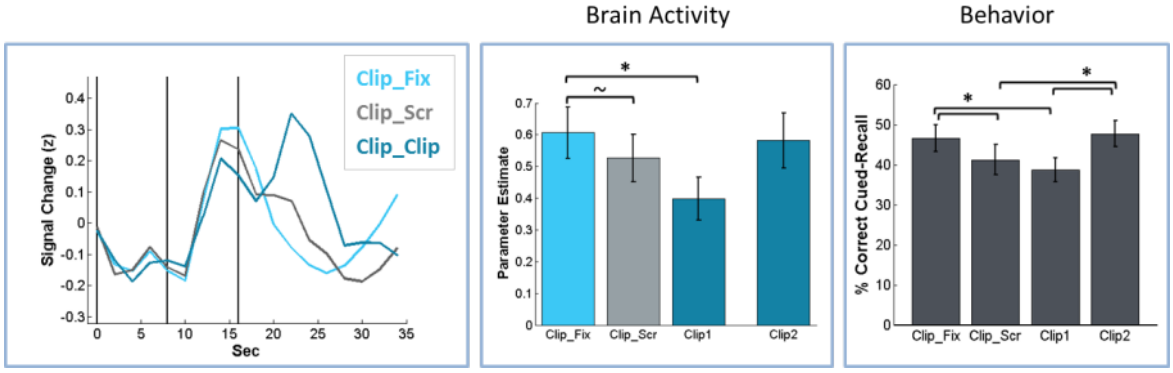


Figure 91: Effect of retroactively-interfering stimuli on hippocampal response at clip offset.

Left panel: average response to single presentation clips (Clip_Fix), clips followed by a visually scrambled clip (Clip_Scr) and a pair of consecutive clips (Clip_Clip) in the right posterior hippocampus. The vertical lines represent the onsets and offsets of the two clips. Middle panel: The amplitude of the response in each condition (beta value estimates), demonstrating a significant effect ($p<0.05$, $\eta p2=0.18$) of presentation condition on the offset-response to the first clip. This was due mainly to a significant reduction in the response to Clip1 of a clip pair relative to a single clip. Right panel: The memory performance in each of the conditions. Adapted from Ben-Yakov, Eshel and Dudai 2013.

The final study addressed the effects of event familiarity on hippocampal processing. We presented the same clips repeatedly, in addition to a set of clips that were presented only once (Ben-Yakov et al., 2014). We found that increased familiarity attenuated the hippocampal response at clip offset, in line with a novelty/encoding signal (Figure 92). In parallel, an onset response emerged in the posterior hippocampus only for familiar clips, in line with a familiarity/retrieval signal (Figure 92). Thus, we observe a temporal dissociation between encoding and retrieval signals linked to a single event, enabling the study of both processes simultaneously. This dissociation, which was not feasible in previous studies that employed brief stimuli, may prove of particular importance in the study of reconsolidation and false memory.

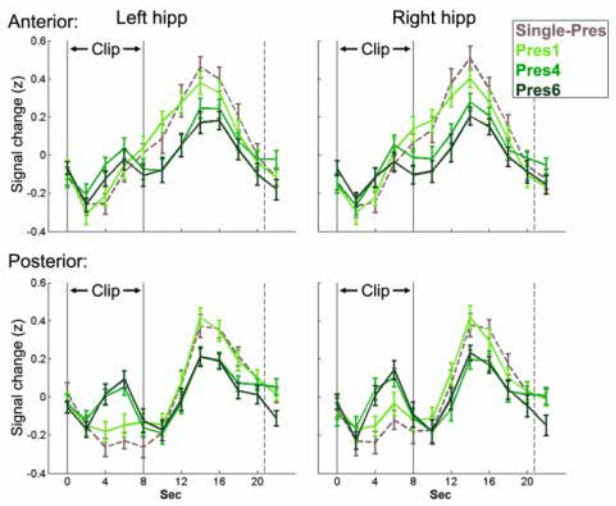


Figure 92: Repetition attenuates offline hippocampal activity and induces an online response.

Mean group BOLD signal (z-scored) during and following single-presentation clips (Single-Pred) and presentations 1,4,6 (Pres1,4,6) of the repeated clips. The black lines indicate the onset (left line) and offset (right line) of clip presentation, while the dashed line indicates the mean onset of the following clip. Error bars represent the standard error of the mean. From Ben-Yakov, Rubinson, and Dudai 2014.

Neocortical pre-encoding involvement in memory formation

It is yet unknown whether natural brain activity preceding the onset of an event plays a role in memory formation. Data from two experiments previously conducted in our lab (Experiment 1 and Experiment 3 in (Ben-Yakov and Dudai, 2011)) were re-analyzed to test this question, and additional experiments performed to clarify the initial data in collaboration with Prof. Rony Paz (from our WP). In these experiments participants encoded audio-visual clips during an fMRI scan. Each clip was preceded by a fixation screen of jittered length. Following the encoding phase, the memory for the clips was assessed using a cued-recall test. In both experiments, the left dorsal anterior insula (peak MNI coordinates: -30, 8, 10; see Figure 93A) demonstrated higher prestimulus activity for subsequently remembered vs. subsequently forgotten clips. Our findings suggest that pre-encoding activity in the left anterior insula plays a role in memory formation. The left insula exhibited a specific pattern of memory-predictive brain activity: deactivation during the movie and activation during the fixation interval (see Figure 93B & Figure 93C). Memory outcome appears to be predicted by the peak activity of the insula after it is has recovered from deactivation. Thus, our results imply that stronger recovery from deactivation promotes the encoding of a novel upcoming event.

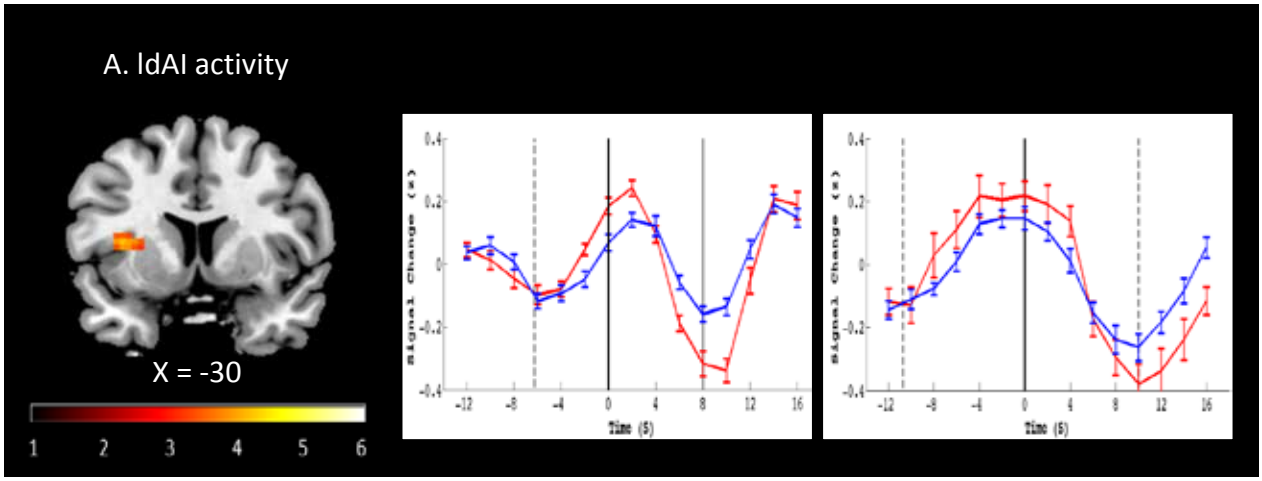


Figure 93

(A) Regions demonstrating higher BOLD pre-encoding activity for remembered clips compared to forgotten clips ($p < 0.001$, cluster size > 38), and mean group BOLD signal (after z scoring each time course) during and following remembered and forgotten clips in Experiment 1 (B) and in Experiment 2 (C). Error bars indicate standard error of the mean (SEM). The vertical lines indicate the onset and offset of clip presentation. lDAI= left dorsal anterior insula. (Cohen, Ben-Yakov, Edelson, Paz and Dudai 2016, under revision).

All in all, our work enabled us to generate a block diagram of the basic cognitive architecture of the initiation of systems consolidation in the human brain. The diagram, coupled to further events in the early stages of systems consolidation as unveiled in the parallel work of Jan Born's group (Tuebingen) in our WP, is presented in Figure 94.

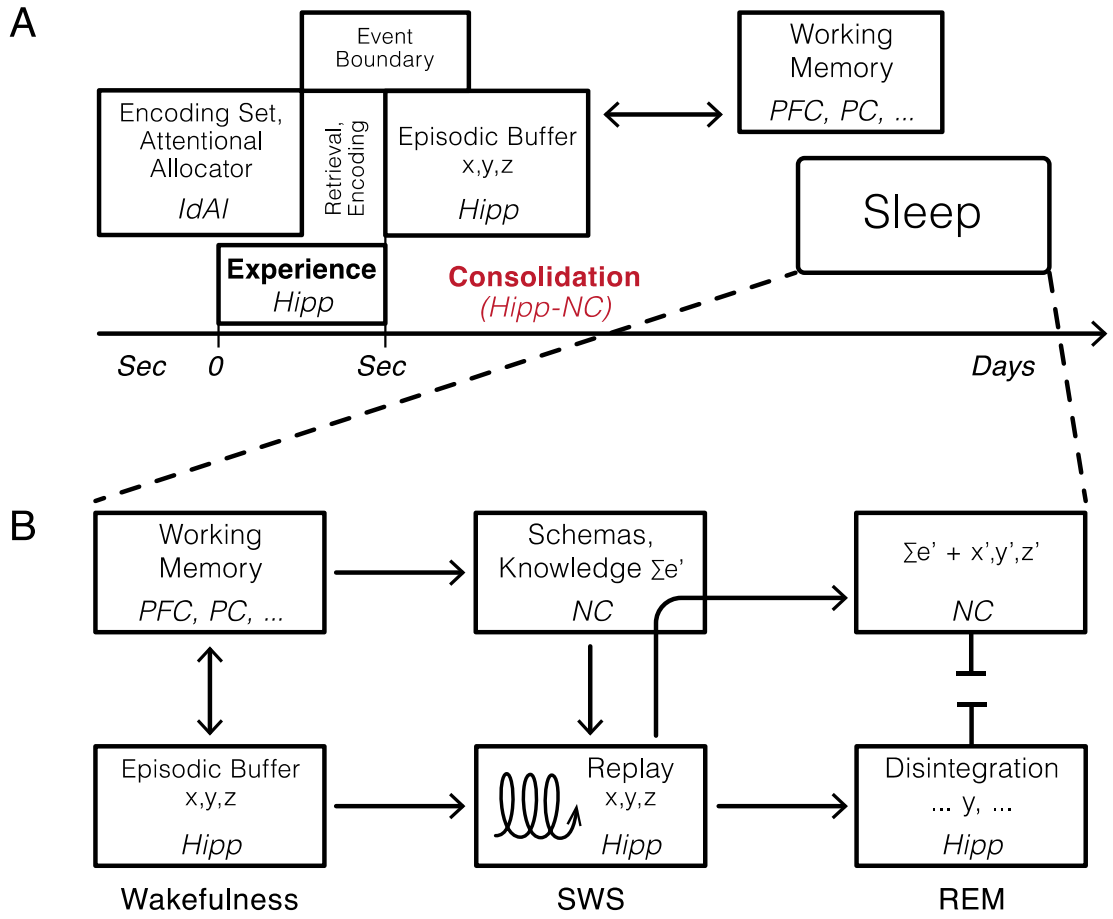


Figure 94: A heuristic block model of the cognitive architecture of selected phases in episodic memory consolidation.

(A). The initiation of consolidation. Activation of an encoding set in the dorsal anterior insula precedes the event to be encoded, which is registered on the fly in the hippocampal system, involving rapid alternations of encoding mode (of the new information) and retrieval mode (of familiar attributes of the experience, encoded in both hippocmapus and neocortex). An automatic episodic buffer, which also suberves working memory related to the ongoing task, is assumed to bind the incoming information into a coherent representation in the hippocampus, the closure of which by a postulated event boundary (computed by modality specific neocortex) sets into action the consolidation cascade. (B). This part of the model is presented to provide a broader picture and is informed by the work performed in Jan Boren's group in Tuebingen in the context of the overall task. Consolidation during sleep. The episodic experiences (X,Y.Z) loading into the hypothetical hippocampal-based buffer is accompanied by EEG theta activity and tagging of memories for reactivation during succeeding sleep. Reactivation that repeatedly occur during slow wave sleep stimulate the passage of the reactivated memory information towards neocortical storage sites where this memory information becomes integrated into pre-existing knowledge networks. Ensuing REM sleep stabilizes the newly formed NC representations via synaptic consolidation and might simultaneously degrade and disintegrate (large parts of) the hippocampal representation. Hipp, hippocampus; IdAI, left dorsal anterior insula; MTL, mediotemporal lobe; NC, neocortex. PFC, prefrontal cortex, SWS, slow wave sleep. For further details on the initial phases in this block model, see text. (Adopted from Dudai, Karni and Born, Neuron 2015.)

The following papers were included in the HBP support period:

1. Ben-Yakov A, Eshel N, Dudai Y (2013) Hippocampal immediate post-stimulus activity in the encoding of consecutive naturalistic episodes. *J. Exp. Psychol:G*, 142, 1255-1263.
2. Ben-Yakov A, Rubinson M, Dudai Y (2014) Shifting gears in hippocampus: Temporal dissociation between familiarity and novelty signatures in a single event. *J Neurosci* 34(39), 12973–12981.

3. Ludmer R, Edelson M, Dudai Y (2014). The Naïve and the Distrustful: State-dependency of hippocampal computations in manipulative memory distortion. Hippocampus 25(2), 240-252
4. Edelson M, Shemesh M, Weizman A, Yariv S, Sharot T, Dudai Y (2015). Opposing Effects of Oxytocin on Overt Compliance and Lasting Changes to Memory. Neuropsychopharmacology. 40(4), 966-73.
5. Cohen N, Pell L, Edelson M, Ben-Yakov A, Pine A, Dudai Y (2015). Peri-encoding predictors of memory encoding and consolidation. *Neurosci Biobehav Rev* 50, 128-142.
6. Dudai Y, Karni A, Born J (2015). The consolidation and transformation of memory. *Neuron* 88, 20-32.
7. Cohen N, Ben-Yakov A, Edelson GM, Paz R, Dudai Y (2016). Prestimulus Activity in the Human Dorsal Anterior Insula Predicts Subsequent Memory (*under revision*).

Talks including data and models involving our WP were delivered by Yadin Dudai in EMBO annual meeting (Heidelberg, October 2014), MIT (The Picower Annual Lecture, May 2015), NYU Advances in Memory Systems Meeting (May 2015), Swiss Memory Workshop, Spiez, Switzerland, Aug 2015), The Friedrich Meischer Lecture (FMI, Basel, Nov 2015), and Brain Mind Institute Memory Workshop (Lausanne, December 2015).

Collaboration: We have collaborated with Prof. Rony Paz from our WP, with Prof. Talma Hendler from SP3 T3.2.4, with Prof. Misha Tsodyks from SP3 T4.3.2 and from SP6, and with Kathinka Evers from SP12.

Data Provenance: The fMRI and behavioural data were collected and analysed by Aya BEN-YAKOV, Neetay ESHEL, Micah RUBINSON and Meytar ZEMER. The data were also analysed by Noga COHEN, at the WIS.

Data deposition: This was conducted and accomplished according to the instructions from HBP. Since there was no HBP data platform prepared by HBP on time, we served as a template for data cards and data repositories, in interaction with the HBP Management.

The data were deposited on a server at:

http://sp3.s3.data.kit.edu/3_3_2/fMRI/Consolidation_of_realistic_episodic_memories_stage_1/
http://sp3.s3.data.kit.edu/3_3_2/fMRI/Consolidation_of_realistic_episodic_memories_stage_2/

http://sp3.s3.data.kit.edu/3_3_2/fMRI/Prestimulus_predictors_of_memory_encoding/

Three **Dataset Information Cards** have been completed (see DICs Task T3.3.2 “Prestimulus predictors of memory encoding”, “Consolidation of realistic episodic memories - stage 1” and “Consolidation of realistic episodic memories - stage 2”).

Brain circuits underlying maladaptive memories in anxiety

Task 3.3.2 Rony Paz (WIS)

Over-generalization of dangerous stimuli is a possible etiological account for anxiety disorders, yet the underlying behavioural and neural origins remain vague. Specifically, it is unclear if this is a choice-behaviour in an unsafe environment ("better-safe-than-sorry"), or also a fundamental change in how the stimulus is perceived. If it is the latter, it means that plasticity mechanisms that occur during learning change the internal representation and affect the memory, so that later recall will be biased. Based on our previous findings showing that aversive learning can indeed affect memory, we hypothesized that such a basic mechanism can contribute to anxiety responses. Specifically, we hypothesized that patients would have compromised perception that contributes to over-generalization. We further hypothesized that this would be paralleled by changes in stimulus representations in sensory-regions that are modulated by affective regions.

We designed a specific behavioural paradigm and show that anxiety-patients have wider generalization for loss-conditioned tone when compared to controls, and do so even in a safe context that requires a different behavioural policy. Moreover, patients over-generalized for gain-conditioned tone as well. Imaging (fMRI) revealed that in anxiety only, activations during conditioning in the dACC and the Putamen were correlated with later over-generalization of loss and gain, respectively; whereas valence distinction in the amygdala and hippocampus during conditioning mediated the difference between loss- and gain-generalization. During generalization itself, neural discrimination based on multivoxel patterns in auditory-cortex and amygdala revealed specific stimulus-related plasticity. Our results suggest that over-generalization in anxiety has perceptual origins and involves affective modulation of stimulus representations in primary cortices and amygdala.

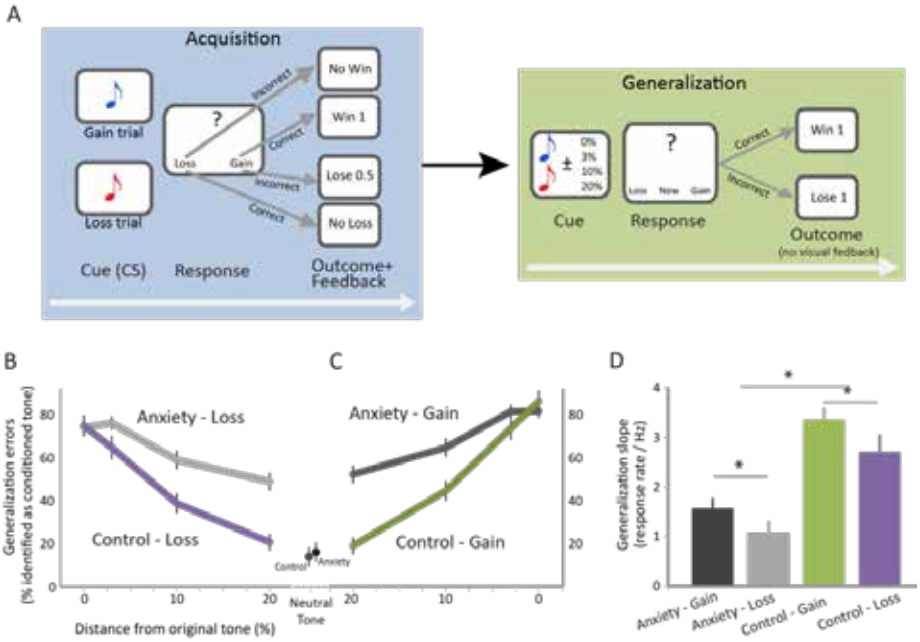


Figure 95: Generalization in anxiety.

(A) Conditioning-phase required subjects to learn to associate one tone (CS-gain) with one button to obtain monetary reward, and another tone (CS-loss) with the other button to avoid monetary loss. In the Generalization-phase that followed (the focus of this study), Subjects heard different tones and had to choose if it is one of the tones that were conditioned in the previous phase (independent if it was the gain- or loss-related tone), or another, new tone. **(B)** Proportion of trials identified as the loss-conditioned-tone, as a function of distance (in % Hz) from the loss-conditioned-tone. **(C)** Same as (B) for the gain-conditioned tone (the graph is reflected only for presentational reasons, the x-axis in both (B) and (C) represents ±3%, ±10%, ±20%). **(D)** Generalization quantified as the slope (averaged over subjects) reveals that there was wider

generalization (smaller slopes) for loss-related tone in both patients and controls, but that patients had even wider generalization. (n=28 GAD, 16 controls).

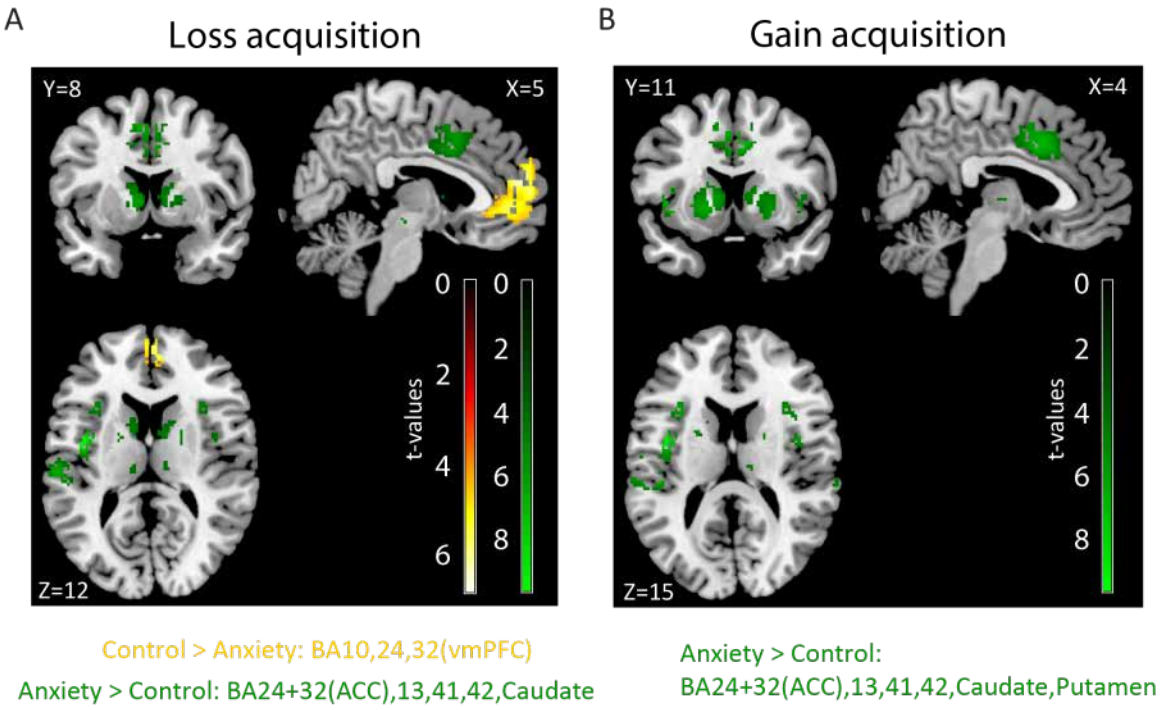


Figure 96: Differential activations during conditioning

(A) During loss-conditioning trials, activations were higher in controls in the vmPFC (yellow) and BA10, and higher in anxiety-patients in dACC, Caudate, insula, and auditory cortices (green). **(B)** During gain-conditioning trials, no regions showed increased activation in controls, but activations were higher in anxiety-patients in dACC, Caudate&Putamen, insula, and auditory cortices (green). Activations are thresholded at p<0.05, FWE SVC, k>10. (n=16 GAD, 16 controls).

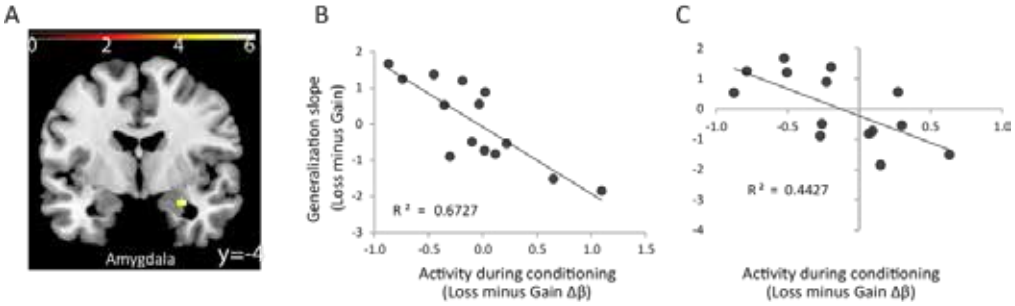


Figure 97: Activations in Amygdala during conditioning are correlated with later Individual generalization difference between loss and gain.

(A) Activation map when using generalization slopes as covariate with a loss>gain contrast in a GAD vs. control model: the amygdala showed significant activation during conditioning that was significantly more correlated with the later individual generalization behavior of GAD patients only. Activations are thresholded at p<0.05, FWE SVC, k>10. **(B)** Correlations between individual activations of anxiety subjects and their behavioral slopes are shown. Notice this correlation plot report the effect size and directionality and do not constitute additional non-independent tests. **(C)** Same as (B), but using activations from anatomical ROIs of the amygdala.

References

Laufer O. Israeli D. Paz R. Behavioral and neural mechanisms of over-generalization in anxiety. Current Biology, In press

Author contributions: R.P., O.L. and D.I. designed the study; O.L. and D.I. conducted the experiments; O.L. analyzed the data; R.P. and O.L. wrote the paper.

Acknowledgments: We thank Dr. Edna Furman-Haran and Nachum Stern for MRI procedures. The work was supported by I-CORE #51/11, ISF #26613, and Minerva-Foundation grants.

Model: Neural mass models of the sleeping brain

Task T3.3.2 - Jan Born (EKUT)

The overarching aim of this project was to model the transformation of hippocampus-dependent memory during sleep-dependent system consolidation. These consolidation processes require an intricate interplay between the hippocampus and other key structures, namely the neocortex and the thalamus. Therefore, as a first step we aimed at modeling brain dynamics during sleep using neural mass models, to better understand the interplay between those structures and elucidate the effect of external non-invasive stimulation techniques on the ongoing brain activity.

Neural mass model of cortical slow oscillations and K-complexes

In stage 1, we developed a neural mass model of the sleeping cortex (Weigenand et al., 2014). This approach allowed for a cost efficient modelling of sleep EEG and provided a detailed understanding of the mechanisms involved in the generation of slow oscillations and K-complexes.

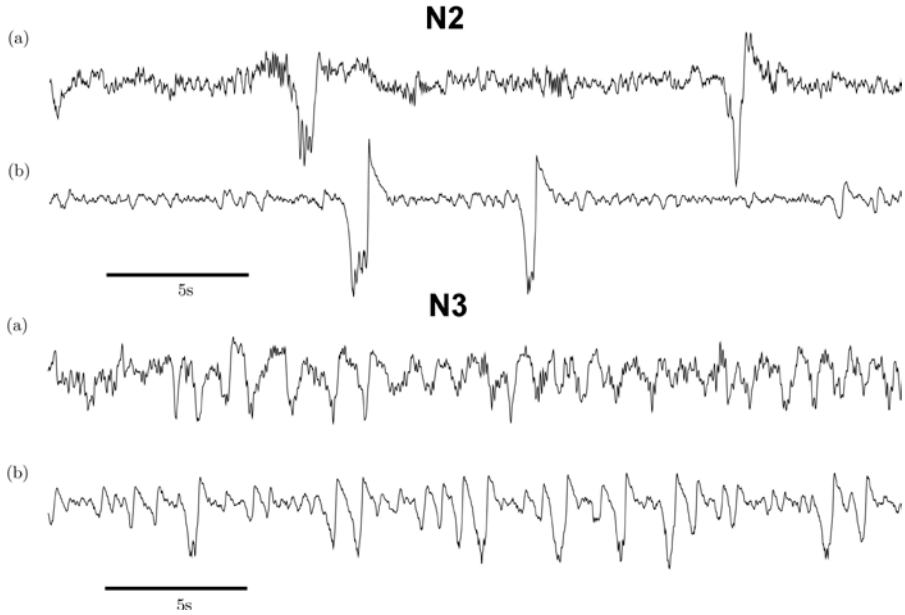


Figure 98: Comparison of model output with human EEG data during NREM sleep.

One epoch of 30s of human EEG data (top) compared with model output. The upper panel depicts sleep stage N2, the lower N3 or slow wave sleep (From Weigenand and Schellenberger Costa et al. 2014).

Based upon our analysis, we could show that on a mesoscopic scale, the cortex is approaching an Hopf bifurcation rather than alternating between two stable states, which contradicts the existing literature of a bistable cortex during slow wave sleep. Furthermore, we were able to show, that the generating mechanisms of K-complexes and slow oscillations differ. While slow oscillations are generated by a plain limit cycle, K-complexes emerge from the interplay between a slow firing rate adaptation and the cortical dynamics, leading to a canard phenomenon.

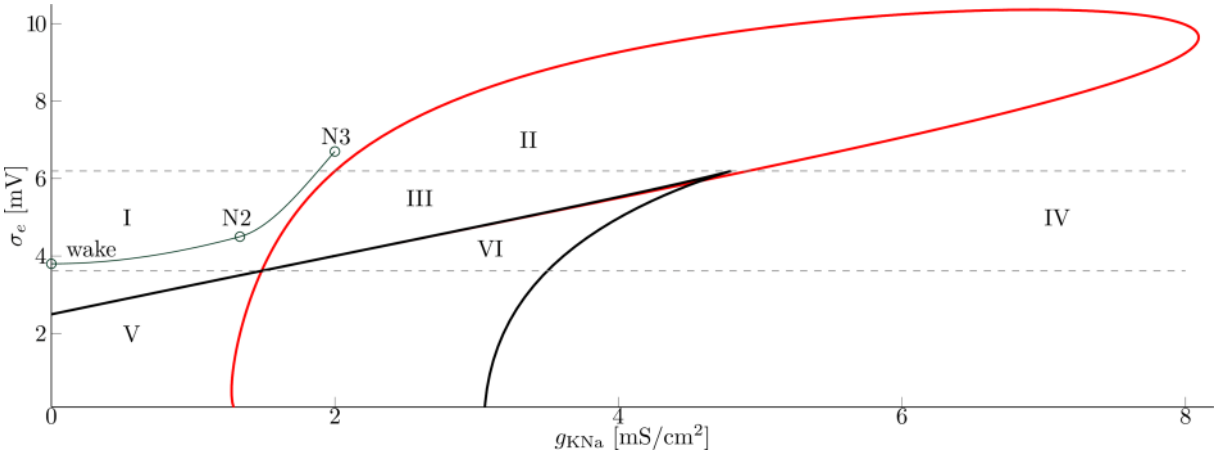


Figure 99: Bifurcation analysis of the cortical neural mass model.

In addition to the previously know bistability, that was generated by two saddle-node bifurcations (black), there is an additional Hopf bifurcation (red) generated by the slow firing rate adaptation included in the cortical model. In the region between the grey lines, there is a canard phenomenon, that generates K-complexes. Above the upper grey line the canard vanishes in a cusp bifurcation and a pure limit cycle remains, leading to large amplitude slow oscillations (From Weigenand and Schellenberger Costa et al. 2014).

Neural mass model of thalamo-cortical interactions and the effect of sensory stimulation.

Hippocampus-dependent memory consolidation during sleep is highly dependent on the interplay between cortical slow oscillations, thalamic spindles and hippocampal sharp wave ripples. Consequently in stage 2 we extended our cortical neural mass model to include a thalamic component (Schellenberger Costa et al., In review). We were able to show, that the inclusion of a T-type calcium and an anomalous rectifier current is sufficient to generate highly realistic spindle oscillations.

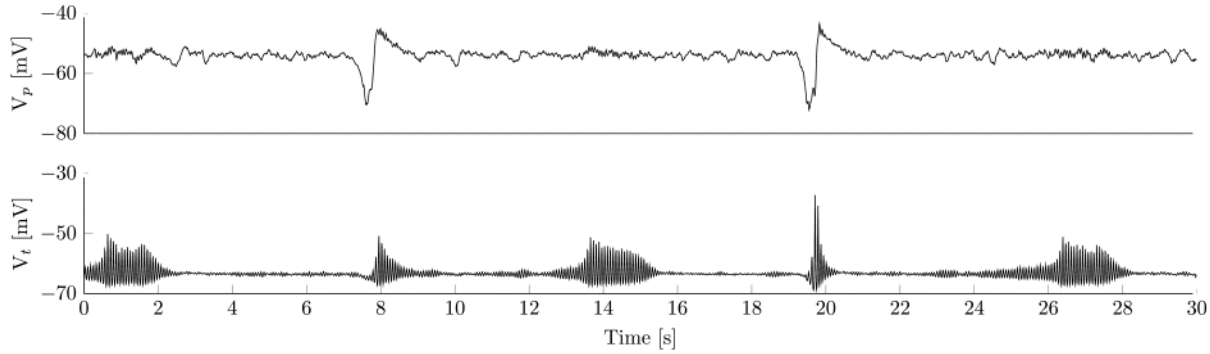


Figure 100: Model output during simulation sleep stage N2.

The upper panel shows the activity of the cortical neural mass, that relates directly to the measured EEG. The lower panel depicts spindle activity on the thalamic relay nuclei. While the thalamic nuclei is able to periodically generate spindle oscillations, there is also a strong coupling between K-complexes/Slow oscillations and thalamic spindles (From (Schellenberger Costa et al., In review)).

To better validate our model and elucidate the effect of external sensory stimulation on the thalamo-cortical dynamics, we reproduced experimental data from a previous stimulation study in humans (Ngo et al., 2013). The model not only reproduces the experimental data to a high degree, but also exhibits the typical grouping of slow oscillatory activity and thalamic spindles.

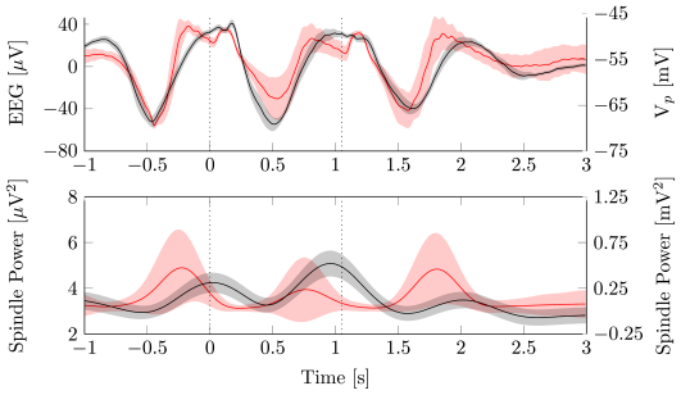


Figure 101: Model output during simulation sleep stage N2.

The upper panel shows the activity of the cortical neural mass, that relates directly to the measured EEG. The lower panel depicts spindle activity on the thalamic relay nuclei. While the thalamic nuclei is able to periodically generate spindle oscillations, there is also a strong coupling between K-complexes/Slow oscillations and thalamic spindles (From (Schellenberger Costa et al., In review)).

Effect of sleep regulation on neural mass models.

Changes in brain dynamics are driven by neuromodulators that are highly specific to the respective brain structure. In our previous work we could show, how certain mechanisms shape the underlying brain dynamics during sleep and link them to neuromodulatory activity. However, to fully understand brain dynamics during sleep one need to understand not only the way those neuromodulators act, but also their intrinsic dynamics. Therefore, we extended our previous approach of the sleeping cortex with a sleep regulatory network, that shapes cortical activity through the release of different neuromodulators (Schellenberger Costa et al., In print).

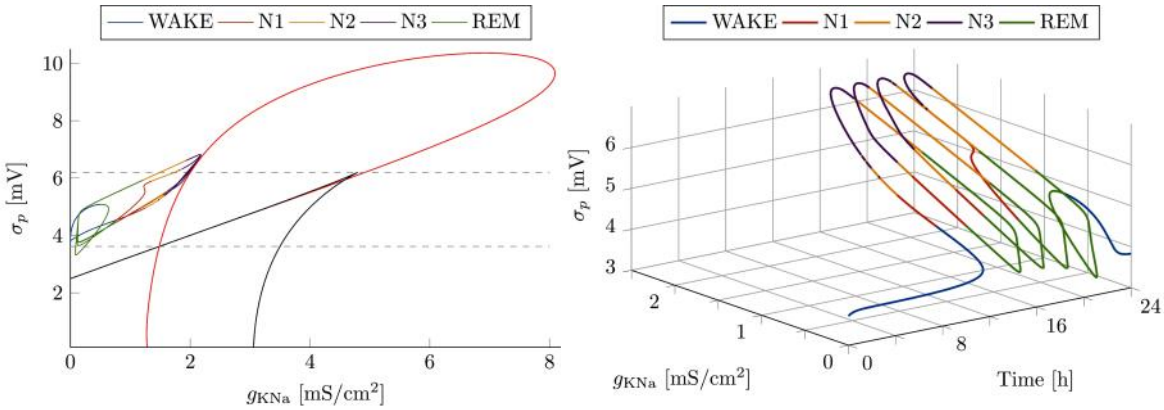


Figure 102: Effect of sleep regulation on a cortical neural mass.

The right panel illustrates the time course of the bifurcation parameters of the cortical neural mass over the course of a full day, exhibiting periods of wakefulness, NREM and REM sleep. The different neuromodulators of the sleep regulatory network (acetylcholine, noradrenalin and extrasynaptic GABA) act on the bifurcation parameters, generating the characteristic REM-NREM cycling observed during human sleep. The left panel depicts the projection of the trajectory into the bifurcation diagram of the cortical model, illustrating, how the different sleep stages depend on the bifurcation structure of the model (From Schellenberger Costa et al. 2016b).

Our approach has the additional advantage, that we are now able to classify the activity of the sleep regulatory network based on EEG activity generated by the cortical neural mass

rather than the intrinsic firing rates of the sleep regulatory network itself, as currently done in the literature. This allows for an easy non-invasive validation of sleep regulatory networks, that was previously not possible. Furthermore, we are able to predict the effect of neuropharmacological interventions on the underlying brain dynamics and therewith the generated EEG.

The following papers were included in the HBP support period:

1. Weigenand A, **Schellenberger Costa M**, Ngo H-VV, Claussen JC, Martinetz T (2014) Characterization of K-Complexes and Slow Wave Activity in a Neural Mass Model, PLoS Computational Biology e1003923, doi: 10.1371/journal.pcbi.1003923,

2. **Schellenberger Costa M**, Weigenand A, Ngo H-VV, Marshall L, **Born J**, Martinetz T, Claussen JC (2015) A thalamocortical neural mass model of the EEG during NREM sleep and its response to auditory stimulation. (In review, PloS Computational Biology)

3. **Schellenberger Costa M**, **Born J**, Claussen JC, Martinetz T (2015) Modeling the effect of sleep regulation on a neural mass model. (In Print, Journal of Computational Neuroscience).

Location of models
The model implementations are publicly available at

[1] Schellenberger Costa, M.: Neural mass model of the cortex during sleep (2014). https://github.com/miscco/NM_Cortex

[2] Schellenberger Costa, M.: Neural mass model of the isolated thalamus during sleep. (2015). https://github.com/miscco/NM_Thalamus

[3] Schellenberger Costa, M.: Neural mass model of the thalamocortical system during sleep. (2015). https://github.com/miscco/NM_TC

[4] Schellenberger Costa, M.: Simulation of the effect of sleep regulation on a cortical neural mass (2015). https://github.com/miscco/ NM_Cortex_SR

The thalamocortical part is complete. Hippocampal components are still missing.

We have a starting collaboration with Maxim Bazhenovs lab and will add our models to the virtual brain project.

3.2 Working Memory

Task T3.3.3 - Lars Nyberg (UMU), Johan Eriksson (UMU)

Review of the cognitive architecture for working memory

Johan Eriksson, Edward K. Vogel, Anders Lansner, Fredrik Bergström, Lars Nyberg "Neurocognitive Architecture of Working Memory", Neuron, Volume 88, Issue 1, p33–46, 7 October 2015

Abstract

A crucial role for working memory in temporary information processing and guidance of complex behaviour has been recognized for many decades. There is emerging consensus that working-memory maintenance results from the interactions among long-term memory representations and basic processes, including attention, that are instantiated as reentrant loops between frontal and posterior cortical areas, as well as sub-cortical structures. The nature of such interactions can account for capacity limitations, lifespan changes, and restricted transfer after working-memory training. Recent data and models indicate that working memory may also be based on synaptic plasticity and that working memory can operate on nonconsciously perceived information.

Data set: Short-term maintenance of conscious and non-conscious information.

Introduction

Working memory maintains information in an easily accessible state over brief periods of time. This feature is required for future goal-directed behaviour and allows us to act beyond the confines of the here and now. As such, working memory is taxed by numerous laboratory and everyday cognitive challenges. Core features of working memory are short-term maintenance of information in the absence of sensory input, distraction resistance, and a connection between the maintained information and prospective actions. Recent research has demonstrated that, contrary to common belief, these features are possible also for information that has been presented non-consciously (Bergström and Eriksson, 2014; Pan et al., 2014; Soto and Silvanto, 2014). Here we further investigate the characteristics of such "non-conscious working memory" by suppressing the conscious experience of a sample stimulus using continuous flash suppression in a delayed match-to-sample task, while measuring brain activity with fMRI.

Materials and Methods

Participants

Thirty participants were recruited from the Umeå University campus area. All participants had normal or corrected to normal vision, right eye- and hand dominance, gave written informed consent, and were paid for participation. Four participants were excluded due to excessive head motion during scanning, misunderstanding instructions, or outlier performance, and the final dataset therefore contains data from 26 participants. All participants gave written informed consent and the study was approved by the local ethics committee.

Stimuli and procedure

Each trial began with an inter-trial-interval (ITI) of 3-9 s before the stimulus presentation. The stimuli consisted of one out of six different grey silhouettes of tools and were presented on a computer monitor An MR-compatible mirror stereoscope was used to isolate the visual input from the left side of the monitor to the participants left eye, and vice versa for the right side. The stimulus to be retained was presented for 3 s, either to both eyes simultaneously (consciously experienced), or only to the non-dominant (left) eye while coloured squares of random composition (mondrians) where flashed with a frequency of 10 Hz to the dominant eye to suppress the stimulus from conscious experience (Tsuchiya and Koch, 2005). During the baseline trials mondrians were presented to the dominant eye while an empty gray background was presented to the non-dominant eye. Critically, the visual experience of baseline and non-conscious trials was the same (experiencing only mondrians).

After a variable (5-15 s) delay period a probe prompted a response that indicated whether the probe matched object identity and spatial position, only object identity, only spatial position, or neither (maximum response time 5 s). The participants were instructed to remember both the object identity and its spatial position. Thus, for the probe to be a "match," it had to be the same object and be in the same spatial position (full match) as the sample. If the probe contained the same object at a different spatial position (object match), different object at the same spatial position (spatial match), or different object at a different spatial position (non-match), it should be answered with a "no match" response. If they had not experienced the target stimulus (i.e., only experienced mondrians) they were instructed to guess on the first alternative that came to mind/gut

feeling (match or no match). Next the participants were prompted to make a detection response to determine if a target stimulus had been presented at all (yes or no; max 5 s). If they had not perceptually experienced a stimulus they were to guess per the same instructions as for the delayed match-to-sample task. Lastly, they estimated their conscious experience of the stimulus on a three-point perceptual awareness scale (PAS; (Sandberg et al., 2010). The participants were instructed and trained to use the PAS scale as follows: 1 = no perceptual experience, 2 = vague perceptual experience, and 3 = clear or almost clear perceptual experience, of the sample stimulus. All responses had an upper time limit of 5 s, after which the experiment automatically continued with next response or trial. Prior to the fMRI experiment all participants were trained on the experimental procedure outside of the fMRI scanner.

The experiment consisted of 192 delayed match-to-sample trials dispersed on three presentation conditions (44 conscious, 108 non-conscious, and 40 baseline [target absent] trials; Figure 103). Since we aimed to focus our analyses on comparisons between hits > baseline and hits > misses there was a larger proportion of full match than non-match trials. Out of the conscious trials there were 20 full match, 8 object match, 8 spatial match, and 8 non-match trials (see below for descriptions). Out of the non-conscious trials there were 78 full match, 10 object match, 10 spatial match, and 10 non-match trials.

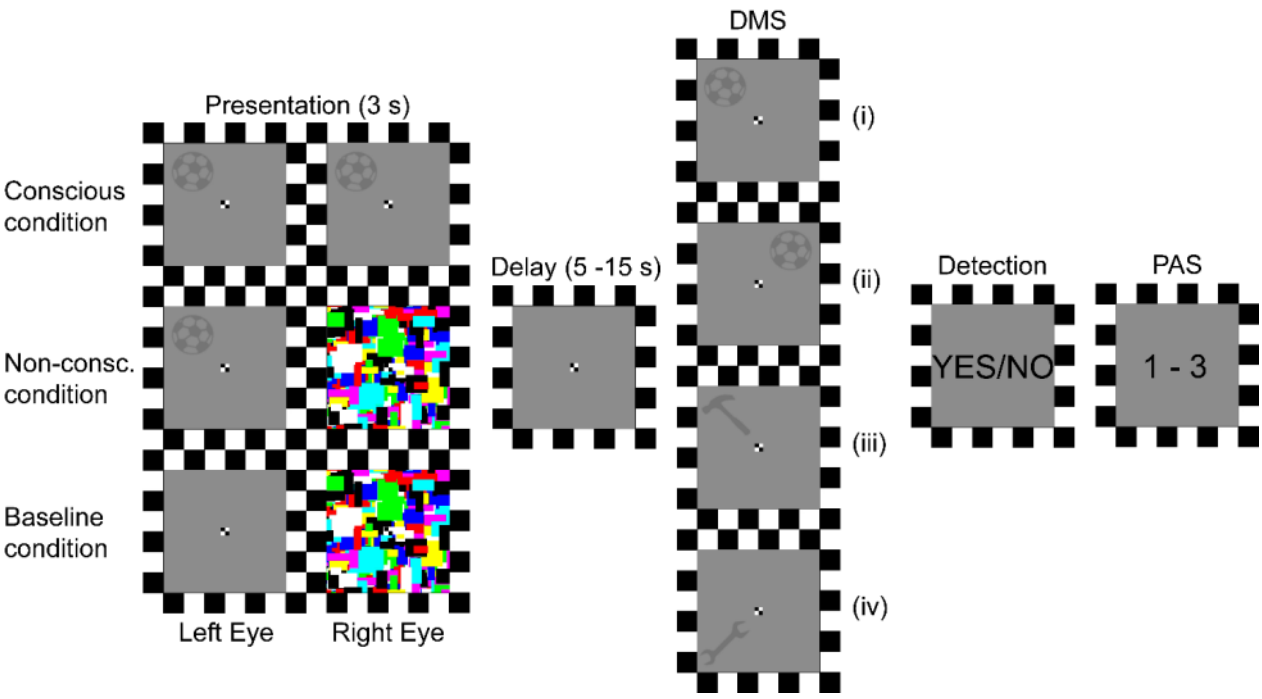


Figure 103: Trial procedure

Depending on the presentation condition, two identical sample stimuli (tools), stimulus and mondrians, or empty background and mondrians, were presented to the left and right eye respectively. The object identity and spatial position of the stimulus was then to be retained for a variable (5-15 s) delay period, until a probe prompted the participants to respond whether or not the probe matched the previously presented position. Next, they responded whether or not a stimulus had been present. Finally, the participants gave an estimate of their perceptual experience of the stimulus. DMS = delayed match-to-sample task; PAS = perceptual awareness scale; (i) = Probe identity and position matches sample; (ii) = Probe identity matches sample; (iii) = Probe position matches sample; (iv) = Probe does not match sample.

fMRI acquisition

The fMRI data were collected at Umeå center for Functional Brain Imaging, Umeå, Sweden, with a GE 3 Tesla Discovery MR750 scanner (32-channel receive-only head coil). Each

participant underwent one session with two functional runs (1230 volumes each) of scanning using a T2∗- weighted gradient echo pulse sequence, echo planar imaging, field of view = 25 cm, matrix size = 96 × 96, slice thickness = 3.4 mm, 37 slices with no inter-slice skip and an ASSET acceleration factor of 2. The volumes covered the whole cerebrum and most of the cerebellum, the acquisition orientation was aligned with the anterior and posterior commissure, and was scanned in interleaved order with TE = 30 ms, TR = 2 s, flip angle = 78◦. Between the two functional runs a high-resolution T1-weighted structural image was collected FSPGR with TE = 3.2 ms, TR = 8.2 ms, TI = 450 ms, and flip angle = 12◦.

Data processing and statistical analysis

Behavioural analysis. Only trials in the baseline and non-conscious presentation conditions with PAS = 1, and trials in the conscious condition with PAS = 3 were used in the statistical analyses, and will for simplicity be referred to as baseline, non-conscious, and conscious trials. Signal detection theory (d′) was used to calculate performance on the delayed match-to-sample discrimination (DMS) and detection tasks (Macmillan and Creelman, 1991). For DMS d′ the signal was defined as the object identity and its spatial position. Hits were therefore defined as a (position and identity) match between sample and probe together with a "match" response, and FAs as a non-match (which includes cases where only position, only identity, or neither was a match) between sample and probe together with a "match" response. For the detection task, hits were defined as the presence of a target stimulus together with a "yes" response, and FA were defined as the absence of a target stimulus (i.e., baseline trials) together with a "yes" response.

Univariate analysis of fMRI data. The software used for processing and analysis of fMRI data was SPM8 (Welcome Trust Centre for Neuroimaging, London, UK), run in Matlab 7.11 (Mathworks, Inc., Sherborn, MA, USA). Before preprocessing, a manual quality inspection using in-house software was done. Preprocessing was done in the following order: slice-timing correction to the first slice using Fourier phase-shift interpolation method, head-motion correction with unwarping of B0 distortions, DARTEL normalization (Ashburner, 2007) using a 12-parameter affine transformation model to MNI anatomical space, and an 8 mm FWHM Gaussian smoothing. DARTEL normalization and smoothing was applied on the contrast images after intra-subject model estimation. For intra-subject modeling a General Linear Model (GLM) with restricted maximum likelihood estimation was used.

The model consisted of the following regressors of interest: Presentation conditions (non-conscious and conscious) by trial epochs (stimulus presentation, delay, and DMS response) by PAS rating (1, 2, or 3) by signal detection category (hits, misses, false alarms, and correct rejections), and presentation condition baseline by trial epochs by PAS rating, and inter-trial intervals (ITI). The model also contained the following nuisance regressors: Missed responses (because of time limit), head motion (six parameters), and physiological noise (six parameters) estimated with temporal variation in white matter and cerebral spinal fluid ((Behzadi et al., 2007). All regressors except for head motion and physiological noise were convolved with the "canonical" hemodynamic response function as implemented in SPM8. The high-pass filter had a cut-off at 128 s, and the autocorrelation model was global AR (1). Model estimations from each individual were taken into second-level random-effects analyses (one-sample t-tests) to account for inter-individual variability. The statistical inferences were made on the whole brain with $p \leq 0.001$ uncorrected for multiple comparisons, $k \geq 20$.

Multi-voxel pattern analysis (MVPA) of fMRI data. Prior to analyzing the data with the Princeton MVPA Toolbox it was preprocessed by correcting for slice-timing and head motion (as described under the univariate analysis). For feature selection on each

individual, only voxels that passed through a binary mask (univariate F-contrast: conscious hits vs. baseline presentations, P <= .0001, uncorrected, k=0) were analyzed. The remaining voxel values were then passed through a high-pass filter (128 s cut-off), and replaced by z-score normalized version. The events of interest were then averaged over time-points. The analyses used a leave-k-out cross validation procedure where k is the number of categories to be classified, and always contains one event from each category. As per (Polyn et al., 2005) a backpropagation-based neural network classifier was used to train and test the patterns in the data, a OnOff-value was calculated as a measure of classification performance, and significance tested using a non-parametric permutation test.

Preliminary Results

The study has not yet been published.

Behavioural performance

DMS d' was significantly greater than zero for consciously (M = 3.85, SE = 0.04, p < .001, one-tailed), but not non-consciously (M = 0.02, SE = 0.04, p = .38, one-tailed) perceived memoranda. Detection d' was significantly greater than zero for consciously (M = 4.05, SE = 0.04, p < .001, one-tailed), but not non-consciously (M = 0.02, SE = 0.04, p = .34, one-tailed) perceived memoranda.

Univariate fMRI results

Compared with baseline (sample absent) trials, there was significant BOLD signal change during the stimuli presentation for consciously perceived samples in visual, parietal, and frontal regions. For non-consciously perceived samples vs. baseline, there was significant BOLD signal change in parts of the visual cortex. During the delay, BOLD signal in inferior occipital cortex was significantly increased during conscious trials. There was no significant difference between non-conscious and baseline trials during the delay. At the delayed match-to-sample response, there was a significant BOLD signal increase for conscious compared with baseline trials in occipital, inferior temporal, pre- and postcentral gyri, and in the cerebellum. For non-conscious vs. baseline trials, there was increased BOLD signal change in the right anterior insula and inferior frontal gyrus.

MVPA results

A characteristic neurocognitive feature of working memory is a sustained response during the delay phase of the task, as was seen during the conscious but not non-conscious trials (above). To further investigate the presence/absence of such sustained response we performed an MVPA analysis (categorizing non-conscious from baseline trials) on the delay period, and also on the presentation and response phases. To validate the soundness of such analyses, we also used the MVPA to categorize conscious from baseline trials.

The MVPA could differentiate between conscious and baseline trials during all trial phases (all p<.0001). For non-conscious vs. baseline trials, the MVPA could differentiate trial type during the presentation (p=.0053) and during the response (p=.017), but not during the delay (p=.66). Critically, the MVPA could also differentiate between non-conscious non-matching and baseline trials during the response (p=.0003).

Discussion

In a delayed match-to-sample task we found significant BOLD signal change for non-conscious trials during the sample presentation and response phase, but not during the delay. A significant signal change during the delay was also absent when using a more sensitive analysis technique (MVPA). Although discrimination performance was non-significant, the BOLD signal differences between non-conscious and baseline trials at the response phase demonstrate that some form of memory trace survived the 5-15 s delay. These findings are consistent with recent research demonstrating that sustained activity during the delay is not necessary for conscious working memory (LaRocque et al., 2013; Lewis-Peacock et al., 2012), and alternative coding mechanisms have been proposed (Mongillo et al., 2008).

The significant signal change at the response during non-conscious trials could in principle reflect simple repetition priming (i.e., the same stimulus in the same location was repeated from sample presentation to response probe during "match" trials). However, such priming is inconsistent with the univariate analysis results demonstrating increased signal change in right prefrontal rather than sensory regions. Critically, the MVPA demonstrated significant categorization performance also for non-conscious delayed non-match-to-sample trials, where a different (non-matching) probe item was shown. Other memory mechanisms than priming are therefore needed to explain the current findings (but see (Marsolek, 2008).
To conclude, we here find neural evidence for sustained memory effects in a delayed match-to-sample task, indicating non-conscious working memory.

A **Dataset Information Card** has been completed (see DIC Task T3.3.3 "Non-conscious short-term memory").

Data Provenance
The data has been, and will be, collected by Fredrik BERGSTRÖM at Umeå Center for Functional Brain Imaging (UFBI), Umeå University, Sweden.

Data Location
Data collection is complete and the data were deposited on a server at http://sp3.s3.data.kit.edu/3_3_3/fMRI_raw_data/

Completeness of data set
The data set is complete and can be used for modeling/simulations.

Data quality and value
The data has been deposited as raw fMRI data (nifti files). As this is one of the first fMRI data sets on non-conscious short-term memory the data are potentially highly valuable to novel models of working memory.

Data usage
The data has not yet been used by anyone outside UFBI.

Publications
The following papers were included in the HBP support period:

1. Bergström F, Eriksson J. 2015. The conjunction of non-consciously perceived object identity and spatial position can be retained during a visual short-term memory task. *Front Psychol* **6**: 1470:1–9.

2. Eriksson J, Vogel EK, Lansner A, Bergström F, Nyberg L. 2015. Neurocognitive Architecture of Working Memory. *Neuron* **88**: 33–46.

Space, Time and Numbers

WP3.4 coordinated by Neil Burgess

WP 3.4 Focussed on the cognitive architecture of spatial cognition. The vast existing literature on the neural mechanisms of spatial cognition was synthesised to find the minimal model capable of explaining the major features of mammalian spatial memory and navigation. We were able to define the major neural representations of location, orientation, environmental structure and movement trajectories in the hippocampal formation and striatal and parietal areas. In addition, we were able to identify the broad nature of the learning rules that must be at play in each area, from analysis of the behaviour of animals under various experimental and neuronal manipulations (e.g. lesions or regional inactivation). Namely Hebbian incidental learning in hippocampal areas and reinforcement learning in striatal areas. Finally, the way the two learning systems should interact, with involvement of parietal and prefrontal areas, was outlined. This synthesis was published in the special issue of Neuron relating to SP3 (Chersi and Burgess, Neuron, 2015).

To validate the proposed cognitive architecture we took two classic experiments on spatial learning and navigation in rodents (Packard and McGaugh, 1996; Pearce et al., 1998) and their analogues in humans (Doeller and Burgess, 2008; Doeller, King, Burgess, 2008) and simulated their behavioral outcomes using the proposed architecture of neural representations, systems and learning rules. This work was successfully completed, showing a good match between behavioural data on spatial navigation in open-fields, with local landmarks, in 'T' mazes, and under inactivation of hippocampal or striatal systems.

This work was in combination with SP4: between WP3.4 and WP4.3 a single post-doctoral researcher, Dr Fabian Chersi, was employed, to work with Neil Burgess, to determine the cognitive architecture of spatial cognition (SP3) and to produce the neural simulation (SP4). The validated architecture and neural simulation will now be applied to a broader set of data on mammalian learning and memory, beyond the purely spatial domain (see Reports on WPs 3.4 and 4.3).

4.1 Identifying and Analysing the Multi-modal Circuits for Spatial Navigation and Spatial Memory

Task T3.4.1 - Neil Burgess (UCL), Fabian Chersi (UCL)

Review of the cognitive architecture of spatial navigation

Fabian Chersi, Neil Burgess "The Cognitive Architecture of Spatial Navigation: Hippocampal and Striatal Contributions", Neuron, Volume 88, Issue 1, p64–77, 7 October 2015

Abstract

Spatial navigation can serve as a model system in cognitive neuroscience, in which specific neural representations, learning rules, and control strategies can be inferred from the vast experimental literature that exists across many species, including humans. Here, we review this literature, focusing on the contributions of hippocampal and striatal systems, and attempt to outline a minimal cognitive architecture that is consistent with the experimental literature and that synthesizes previous related computational modeling. The resulting architecture includes striatal reinforcement learning based on egocentric representations of sensory states and actions, incidental Hebbian association of sensory information with allocentric state representations in the hippocampus, and arbitration of the outputs of both systems based on confidence/uncertainty in medial prefrontal cortex. We discuss the relationship between this architecture and learning in model-free and model-based systems, episodic memory, imagery, and planning, including some open questions and directions for further experiments.

Model of spatial navigation and spatial memory

Introduction

Spatial navigation, although one of the most common actions for humans and animals, is a complex task that involves the processing of a variety of sensory and proprioceptive stimuli (e.g. visual, vestibular and motor information), the storage and recall of memories about location and events, and the elaboration of plans.

There are two main mechanisms utilized in spatial navigation. The first one, referred to as a "response strategy" and mainly implemented in the striatum, relies on following well-learned state-response associations. The second one, referred to as a "place strategy" and implemented mainly in the hippocampus, utilizes flexible internal representations of the spatial layout.

Model details

Visual system

The simulated rat has been endowed with a simple visual system that allows it to acquire two types of information about the environment: the colour of the observed objects and their distance (from the observer). In the current implementation the visual field extends from -160 to +160 degrees, and is subdivided into small regions (see Figure 104), each one assigned to one neuron. One of the characteristics of this encoding scheme is that neurons are mostly silent, with only a few of them firing at a high rate when an object enters its receptive field.

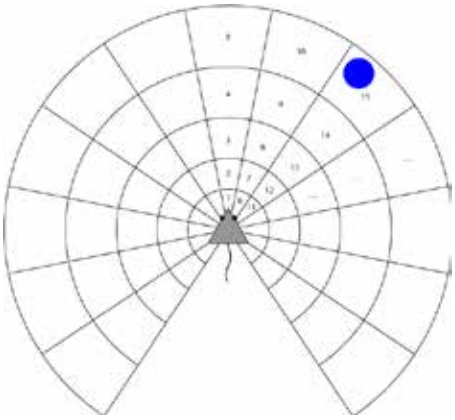


Figure 104: Schematic representation of the neural representation of the rat's view field.

Each region is associated to a specific neuron (numbered). When an object, e.g. the landmark, is spotted, neurons corresponding to the interested regions become active.

The hippocampal circuit

The hippocampus has here been implemented as single layer of firing rate-based neurons with no lateral connections. For sake of simplicity we assume that its neurons encode the position of the agent in an absolute reference frame. The typical response function of place cells is a Gaussian-like activity profile centered on the location it represents. In the current model, we have implemented a mechanism that generates a new place cell whenever the distance to the closest field center is higher than a specific value.

When the rat encounters a salient object (i.e. the hidden platform or food reward in the experiments considered here) the connections between the place cells and a 'goal cell' (Burgess and O'Keefe, 1996)are strengthened according to the following Hebbian-like learning rule:

$$\Delta w_{ij} = \eta \cdot v_i^{(HPC)} \cdot v_j^{(G)}$$

where w_{ij} is the connection weight between neuron i and neuron j, η is the learning rate, $v_i^{(HPC)}$ and $v_j^{(G)}$ are the firing rates of the *i-th* neuron in the hippocampus and of the *j-th* goal neuron, respectively.

The end result is the formation of an object-specific "value function" that has it highest value centered on the location in the environment where the object (i.e. the platform) is. This surface can be used to guide behavior. In particular, given the value function and the current coordinates of the rat, it can test adjacent positions in order to determine the direction in which value increases and use this information to trace a path that leads to the object.

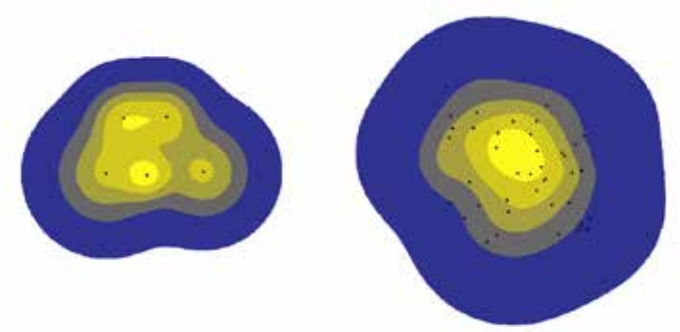


Figure 105: "Value function" that encodes the position of a specific object through the superposition of multiple receptive fields.

The striatal circuit

The stream of transformations that convert the sensory input into an action is shown in Figure 106.

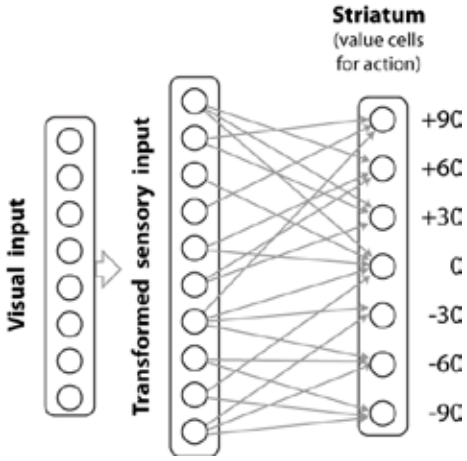


Figure 106: Striatal circuit employed in this architecture that assigns (a value of) an action to each sensory input.

Neurons in the transformed sensory layer project to neurons in the dorsal striatum in an all-to-all manner. The latter neurons represent the values of the possible actions associated with any given sensory state playing here the role of "critics" for the state-action associations.

The sensory and striatal neurons can be thought of as representing state-action combinations for reinforcement learning, thus their connection weights can be learned by means of the Q-learning rule (Watkins, 1989), as follows:

$$Q_a = \nu_a^{(STR)} = F\left(\sum_{i=1}^{N} \nu_i^{(Sens)} \cdot w_{i,a}\right)$$

where Q_a is the expected discounted return obtained by performing action a (in our case one of the 12 angles of rotation) in the current state, $\nu_a^{(STR)}$ is the firing rate of the striatum neuron corresponding to action *a*, *F* is the response function of the neurons, *N* is the total number of sensory neurons, $\nu_i^{(Sens)}$ is the firing rate of the active sensory neuron, and $w_{i,a}$ is the weight between the sensory neuron and the striatum neuron.

At every time step the rat can use the available information about the current and the past *Q* values and the occurrence of a reward to update its internal model by means of the standard Q-learning equations:

$$\Delta Q_{s\text{-}1,a\text{-}1} = \eta \cdot [\, R + \gamma \cdot max_{a'}\, (Q_{s,a'}) - Q_{s\text{-}1,a}\,]$$

$$\Delta w_{i,a\text{-}1} = \Delta Q_{s\text{-}1,a\text{-}1} \cdot \nu_i^{(Sens)} \cdot \left(\sum_{j=1}^{N} \nu_j^{(Sens)}\right)^{-1}$$

where $Q_{s\text{-}1,a\text{-}1}$ is the Q value of action *a-1* in state *s-1*, η is the learning rate, *R* is the reward, γ is the discount factor, and $max_{a'}\, (Q_{s,a'})$ is the maximum *Q* value that can be reached from the current states computed on all possible actions *a'*.

Note that the striatum does not receive information about the head direction, so the sensory vector is aligned to the heading direction and not to a global reference direction.

Results

The above described architecture has been used to control a simulated rat in a variant of the Morris Water Maze described by (Pearce et al., 1998). In this set up, in contrast to the original experiment, a landmark indicates the close-by location of a submerged platform. At the start of each trial, the rat is placed in a random location far away from the platform. Every 4th trial the platform together with the landmark are randomly moved to one of 8 locations. As the rat learns more about how to reach the submerged platform, its behavior shifts from random to goal directed and then to habitual, utilizing the hippocampus and then the striatum respectively to make reasoned decisions.

In the current implementation, every 250 ms a decision procedure is called which selects a new movement direction. Five percent of the times this decision is completely random. This ensures that the rat does not get trapped in cyclical movements. In the remaining 95% of the times, utilizing the mechanisms described above the hippocampus and the striatum each produce an output vector that contains the “values” (encoded as neuronal firing rates) of 12 directions in which the rat can move. The final decision about the movement is taken by a winner-take-all mechanism, possibly implemented by the prefrontal cortex, which selects the most “valuable” one (i.e. the one that most rapidly leads to the goal location).

The left panel in Figure 107 shows the average performance of 4 groups of 30 simulated rats over the course of 11 sessions, each composed of a set of 4 similar trials. Following the original experiment we have deactivated the hippocampus of 2 of the 4 four groups of animals. The corresponding results are indicated by the curves with the full dots. The upper curve represents the performance on the first trial of each set, while the lower curve represents the performance on the fourth trial. As can be seen the performance of intact animals on the first trial is worse. This is because the hippocampus drives the rat to the location where the platform was situated previously, as opposed to the location indicated by the landmark. In a few trials the rat is able to learn the new platform location. Interestingly, also rats with damaged hippocampus are able to solve the task using only procedural memory, and reach the hidden platform in a time that is shorter than the control rats on their first trial!

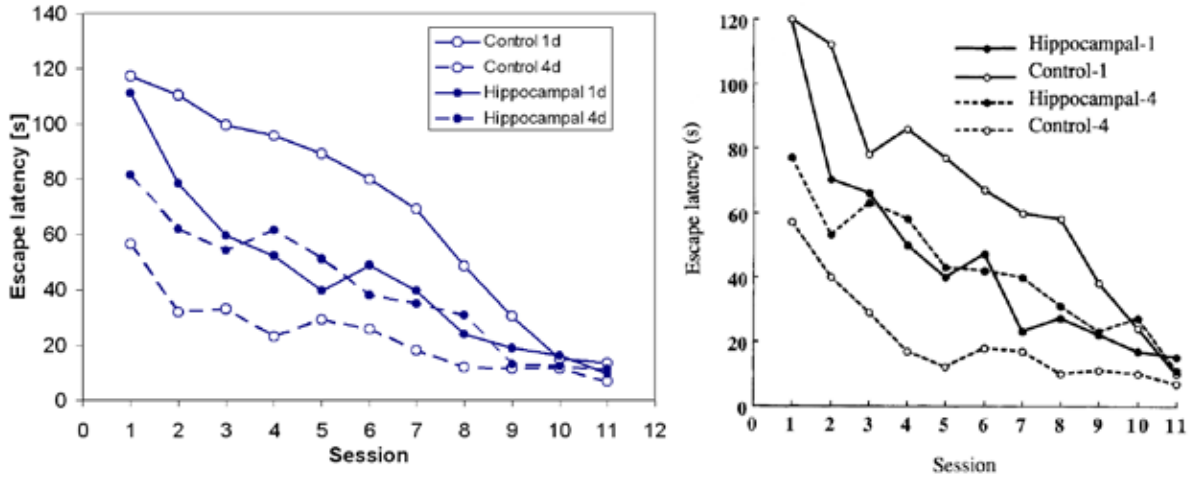


Figure 107

Left panel: average performance of the rats expressed as the time required to reach the escape platform. Right panel: performance of real rats as recorded by Pearce et al. (1998). Empty points represent the behavior of intact control rats, while full points represent animals with lesioned hippocampus.

In a second experiment we simulated the task of (Packard and McGaugh, 1996), in which rats are placed in a cross maze and trained to reach the end of a baited arm. In the test phase, rats were place in the arm opposite to the training home position and monitored as they find their way to the food location.

The left panel of Figure 108 represents the percentage of choice type, whether based on a place strategy or on a response strategy, of all control rats on day 1, on day 8 and on day 16. As can be seen, as the rats spend more time in the maze solving the tasks they switch from a goal-directed strategy to a stimulus-driven one, indicating the formation of habits. As a comparison, in the right panel of Figure 108 we report statistics on the behavior of rats lesioned either in the caudate nucleus on in the hippocampus as found in the paper mentioned above.

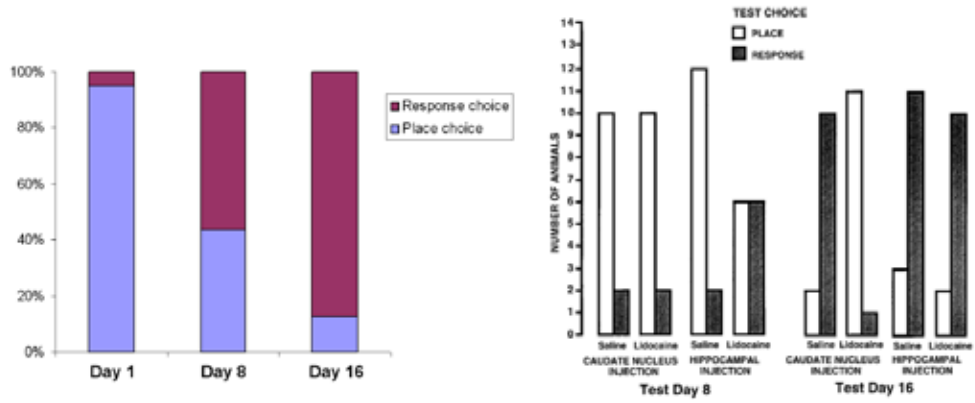


Figure 108

Left panel: distribution of strategy choices (response or place strategy) during session 1, session 6 and session 11. The last two approximately correspond to Test Day 8 and Test Day 16 in the graph in the right panel reporting the results of the experiment conducted by Packard and McGough (1996).

Conclusion

By examining the existing literature, we identified a simplified cognitive architecture of spatial navigation (Chersi and Burgess, 2015). We have implemented this architecture at the level of firing rate neurons and Hebbian and Reinforcement learning rules for synaptic plasticity and instantiated it in a simulated agent. Here we have validated the model by demonstrating that it is capable of reproducing some of the classic experimental tests of spatial navigation. Our next steps will be to see how this same cognitive architecture generalizes to classic non-spatial tests of learning and planning (e.g. (Daw et al., 2011)).

A **Dataset Information Card** has been completed (see DIC Task T3.4.1 "Rat navigation simulation").

Provenance of the data:

The data we utilized in our work was taken from 2 studies published by (Packard and McGaugh, 1996) and (Pearce et al., 1998).

Data Location:

The data we produced is located at http://se/data/kit/edu/SP3/3.4.1/Simulation Results

Self-analysis of the value and completeness of data:

The utilized data is of great value and complete. The data produced by us is perfectly in line with the existing data.

Indication of who has used this data so far and for what

As far as we know, nobody has utilized our data. On the other hand partners from the TU Munich are using the model we have implemented.

List of publications:

Chersi F., Burgess N., 2015. The cognitive architecture of spatial navigation: Hippocampal and Striatal contributions. Neuron 88: 64-77.

Collaborations and interactions with other partners:

University of Leeds (Marc de Kamps)

University of Manchester

University of Munich

EITN Paris

Capabilities Characteristics of the Human Brain

WP 3.6, coordinated by Stanislas Dehaene

The aim of WP 3.6 was to investigate some capabilities that seem either unique, or massively more developed, in the human brain compared to the brains of other non-human primates.

Indeed, any future HBP simulation that claims to capture the functioning of the human brain will have to provide an answer to the issue of the origins of the remarkable capabilities of the human species for high-level cognition. Furthermore, an understanding of human-specific traits is probably crucial to progress in understanding diseases such as aphasia, dyslexia, autism or schizophrenia, which all comprise a perturbation of high-level representations that will be very hard to model in non-human species.

During the Ramp-Up Phase, we decided to focus on three aspects that may hold the key to human specificity.

1. Symbolic thought. Thomas Hannagan and Stanislas Dehaene reviewed how letter and number symbols are represented in the human brain, and provided a model of their acquisition and a distinct model for how vectors of neural activity can implement number symbols.

2. Linguistic and non-linguistic sequences. Christophe Pallier, with Muriel Fabre, Florent Meyniel, Liping Wang and Stanislas Dehaene, investigated how sequences of words or just sounds are represented in humans, and what minimal properties of sequences suffice to induce human-unique activations of language areas.

3. Social brain and theory of mind. Riitta Hari and Lauri Parkkonen investigated the brain networks involved in various levels of social communication, from eye gaze and body movement to theory-of-mind, and they studied a novel social paradigm involving two-person interaction during MEG scanning.

5.1 Symbols and their manipulation

Task T3.6.1 - Thomas Hannagan (CEA) and Stanislas Dehaene

Introduction

The brain is capable of attaching internal symbols to abstract mental representations (e.g. number 3), and of operating on those symbols to perform meaningful computations (e.g. 2+1=3). This task investigated what is known about this symbolic ability, and how it can be formalized. In the review part, we examined what is known about the emergence of specialized areas for the symbols of letters and numbers in the brain. This work has led us to formulate two hypotheses, one of which was computationally investigated, as described in the second part. Finally, we developed a review and model of how numerosity information is represented in the brain by an internal "vector code" that supports a successor function, capable of moving from a neural representation of number *n* to a representation of *n+1* (this work is primarily based on the work of Andreas Nieder, who was initially responsible for T3.6.1 but resigned).

Review of the cognitive architecture for the emergence of symbol-related areas

Relevant paper: Hannagan T, Amedi A, Cohen L, Dehaene-Lambertz G, Dehaene S (2015). "Origins of the specialization for letters and numbers in ventral occipitotemporal cortex." *Trends in Cognitive Sciences*, 19(7), 374-382.

Deep in the occipitotemporal cortex lie two functional regions, the visual word form area (VWFA, (Cohen et al., 2000)) and the number form area (NFA, (Shum et al., 2013)), which are thought to play a special role in letter and number recognition, respectively. The VWFA (Figure 109, central panel, orange area) is a small area downstream of the ventral visual system, in the left lateral occipitotemporal sulcus (OTS), which is consistently activated by visual letters and words (Cohen et al., 2000). Due to fMRI signal dropout around its location, the NFA (Figure 109, central panel, green areas) was discovered only 3 years ago by electrophysiology: it is involved in representing number symbols and is situated a bit more lateral and anterior than the VWFA (Shum et al., 2013). The VWFA and the NFA are puzzling for a number of reasons. First letters and numbers are very similar stimuli from the standpoint of image statistics. If, as is often argued, specialized form areas in vOTC are only the result of a self-organizing phenomon that is guided by the similarity in visual features, why then should areas for letters and numbers ever be dissociated? Why not a single symbol form area for both types of symbols? More puzzling still: how to account for the fact that such areas both exist at the same location in congenitally blind subjects, and that they are activated for symbol images that have been turned into sounds (the so-called "soundscapes", (Abboud et al., 2015)), although they are *not* activated for spoken words? Our work focused on making sense of the origins of these symbol form areas (Hannagan et al., 2015).

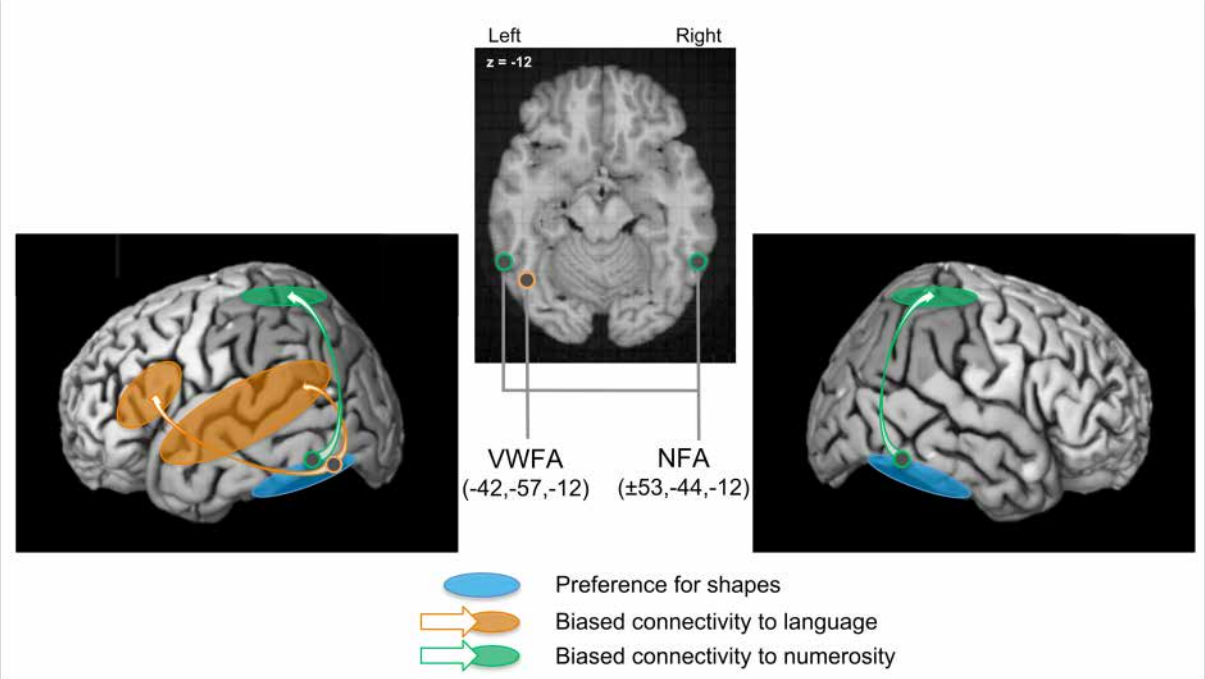


Figure 109: Two converging hypotheses for the origins of symbol form areas: biased connectivity (white arrows) and a preference for shapes (blue areas).

(Left) Left lateral view. (Middle) Axial cut (z = -12). (Right) Right lateral view. Arrows and areas are schematic and only indicative.

We proposed two hypotheses to make sense of the puzzle of symbol form areas (Figure 109). The "biased connectivity" hypothesis holds that form areas emerge at cortical sites that exhibit a higher density of white-matter fiber tracts to and from the cortical circuits that are crucial for the target task (Figure 109, white arrows). This would explain why form areas are conserved at the same coordinates in normal sighted readers and in blind subjects. According to this view, where a symbol form area should mature in the cortex is not so much determined by the visual properties of the stimulus as by some pre-existing (possibly innate) structural connections at this site with the regions that are targeted by the symbol.

The “shape hypothesis” holds that the intrinsic circuitry of the inferotemporal cortex makes neurons in the VWFA and the NFA particularly apt at recognizing the shapes of objects (Figure 109, blue regions). The shape hypothesis can explain why images, soundscapes, and Braille preferentially activate the VWFA and the NFA, whereas spoken words selectively activate the Auditory Word Form Area (DeWitt and Rauschecker, 2012) but not the VWFA (Cohen et al., 2004). According to this view, the differences in selectivity between these areas are not due to an innate lack of inputs from other sensory channels, but to the specific tuning of neurons in the VWFA and NFA to geometrical shape features.

We suggested that both hypotheses were needed to explain the data. While several sectors of the cortex, such as the vOTC, may be tuned to invariant shapes, other cortical sectors may exhibit a stronger connectivity to language- or number-related regions, and symbol form areas would always emerge at the intersection of these two sectors.

This work was presented in the SP3 highlight talk of the 2015 HBP summit. Some of the predictions listed in our review are currently being tested. Two recent experiments at Unicog have endeavoured to test the biased connectivity hypothesis, with functional

connectivity in neonates, and with structural connectivity in children before and after first grade (Moulton et al., submitted). These experiments are now at the stage of data analysis: the analyses so far have neither supported nor invalidated the hypothesis.

We also set forth a number of requirements for a computational model of the emergence of symbol form areas. A unified model should: (1) operate at a human level of performance on letter and number recognition, (2) account for the reproducible location of symbol form areas in absolute Talairach coordinates and relative to other areas, (3) account for the metamodality of symbol form areas, and (4) display the developmental trajectory observed in humans. Such a model may require combining multi-sensory deep convolutional networks with self-organizing models, in a way that is informed by the target systems of language and numerosity that feedback to vOTC.

Exploring the shape hypothesis: mirror invariance in convolutional networks.

Reference paper: Hannagan T, Eickenberg M, Yazdambaksh A, Léveillé J, Pegado F, Grainger, J. (in revision). Mirror invariance in the visual system: a deep learning exploration. PLoS Comput Biol.

The shape hypothesis holds that neurons across the vOTC are invariant to useful, commonly encountered geometric transforms, such as mirror reversals. We thus studied mirror invariance in the OverFeat network, which is a state-of-the-art convolutional network trained on a large-scale dataset of real-life images (Sermanet et al., 2014). We found that the upper layers of OverFeat exhibit mirror invariance for faces, tools, and houses, and that despite having never been trained to categorize stimuli into letters, this mirror invariance for generic objects also automatically transfers to letters (Hannagan et al.,in revision). The finding that OverFeat shows mirror generalization for letters is in line with the shape hypothesis. It is also consistent with the behavior of children who are learning to read (Lachmann, 2002) and supports a perceptual (as opposed to motor) view of mirror errors in reading (Brennan, 2012).

A layer-by-layer analysis revealed that mirror invariance does not culminate in the output layer, but rather, in an intermediate processing stage of the network (layer 7). In an effort to bridge the interpretation gap between neural network activations and cortical activations, we collaborated with the group of Bertrand Thirion (SP2), who had built a predictive mapping between activities in the OverFeat network and in the cortex (Eickenberg et al., submitted).

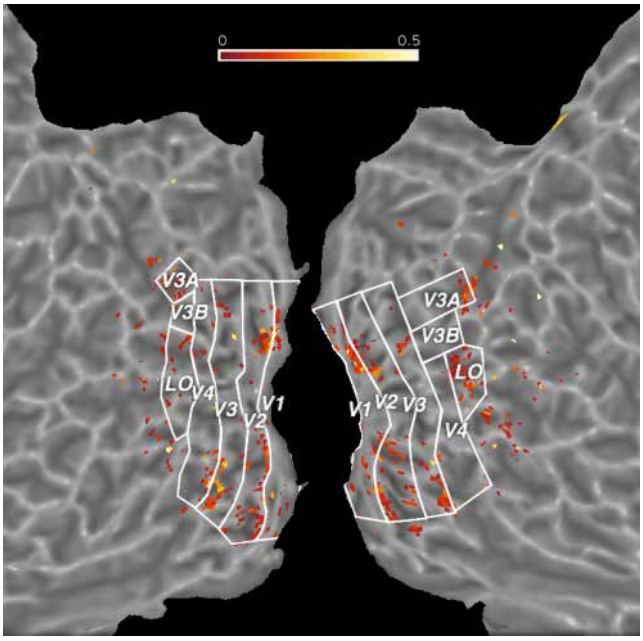


Figure 110: Regions of mirror invariance as predicted by layer 7 of OverFeat and projected onto a flat cortical map, collapsed across six categories of interest: faces, tools, houses, letters, pseudofonts and pseudowords (Hannagan et al., in revision).

Mirror invariance is sparse, bilateral, includes early visual areas, and cuts across the ventral and dorsal streams, along the inferior/superior axis. In higher visual cortex, a relatively large number of contiguous mirror invariant voxels stand out in areas LO and V3A (especially right lateralized). Notice the genuine prediction that mirror invariant voxels should accumulate along at the V1/V2 and V3/V4 borders, which have mirror retinotopy.

The predicted cortical correlates of mirror invariance, shown in Figure 110, were consistent with experimental fMRI data in human adults, as well as electrophysiology in the macaque. But unexpectedly, the mapping also predicted that mirror invariance should surge at the frontier between V1 and V2, and between V3 and V4 (but *not* at the frontier between V2 and V3). With hindsight, it is obvious that retinotopic maps are left-right mirror inverted between these areas: voxels lying astride these borders are thus likely to respond similarly to the presentation of an image and its mirror.

To summarize, we have established that mirror invariance initially transfers from generic objects to visual symbols in convolutional networks of the ventral visual system, and have helped better characterize the mechanisms behind this phenomenon. This work is also a demonstration that network-to-cortex mappings at the level of voxel activations are informative and worth investigating further. This work was conducted in collaboration with Michael Eickenberg, then a graduate student in the group of Bertrand Thirion (SP2) at Neurospin. The OverFeat model is available in the machine learning library "Scikit-Learn Theano", which is hosted and curated by the Parietal research team at Neurospin. The network-to-cortical mapping is not yet published, but will be deposited at the same location after publication.

- ⇨ A dataset containing the cortical simulations from OverFeat for different geometric transforms of faces, tools, houses, letters, strings of letters and strings of pseudoletters, has been deposited at the following address: http://s3.data.kit.edu/3.6.1

A model for the emergence of the number sense

Reference paper: Hannagan T, Dehaene S (in preparation) A random-matrix theory of the number sense.

This last section considers how we access the meaning of symbols, focusing on the numerosity system. What is the origin of our ability to represent numbers? The evidence is now overwhelming for an approximate number system shared between several species, which relies on number neurons (Nieder, 2013): cells whose activity varies in a systematic way for a given number of objects or events, independently of low-level properties (e.g. size, spacing, intensity). Existing computational models (Dehaene and Changeux, 1993; Stoianov and Zorzi, 2012; Verguts and Fias, 2004) have mostly focused on explaining the average tuning curves of number neurons as well as the neural wiring and/or learning rules that could give rise to them. But these models are challenged by one strong experimental result: the approximate number sense does not appear to be learned, being already detectable in neonates of 36 hours of age (Izard et al., 2009). There is currently no model of how a cerebral representation of number could possibly exist in absence of training or detailed handwiring.

We have devised a new model to explain the origins of the approximate number system (Hannagan et al., in preparation). Our starting point was to consider number states rather than number neurons: numbers are encoded by a vector of firing rates over a population of neurons, and successive numbers are obtained by applying to this vector successive powers of a random matrix. This random matrix model can be seen as a cortical implementation of the von Mises eigenvector algorithm (Mises and Pollaczek-Geiringer, 1929), and radically breaks with previous computational accounts of number neurons. However, it can explain the observed logarithmic compression in the average neural tuning curves in a principled way (consistent with the Weber-Fechner law; (Shepard et al., 1975), and our current analyses suggest that it can also reproduce the detailed tuning curves of number neurons. Critically, those results are obtained without training, thus confronting heads-on the innateness of the approximate number system.

Here we present a direct comparison of the model to the electrophysiological data. Nieder and Merten conducted a detailed electrophysiological study of the properties of PFC number neurons for a large range of numerosities, testing monkeys with numbers up to 30 (Nieder and Merten, 2007). Figure 111 compares the activation of the model's units to the empirically observed average firing rates. It can be seen that the model behaves quite similarly to the data collected during the sample period in Nieder and Merten's delayed match-to-sample experiment.

The first row in Figure 111 shows the normalized firing rates for all numbers (x-axes) of all the recorded neurons (y-axes) that were found to be significantly selective to number, in the sample period (150 cells) and delay period (138 cells). These firing rates are compared to the activity of all non-zero units in the model (1000 units). All neural firing rates and unit activities were rescaled row by row, i.e. we divided all the responses of each neuron by its largest response, yielding normalized responses between 0 and 1.The "white crest" of maximal activation allows one to appreciate at a glance how units are distributed for number preference. The fact that yellow regions widen around the white crest as numerosity increases, also reflects the increasing bandwidth of tuning curves. Finally, a line-by-line inspection of these graphs suggests a much richer variety of tuning curves for number neurons than is usually assumed. In particular, some units appear to saturate rapidly for numerosity, while others show disconnected "islands" of high activity, suggesting selectivity for multiple numbers. A conservative statistical analysis revealed that about 10% of units in the experimental data were selective for multiple neurons, against 8.3% for the model.

The second row in Figure 111 proposes a visualization of the number lines that are implied by number states, as observed in the data and in the model. Number states are projected onto the plane using multidimensional scaling, and interpolated by a parameterized polynomial. This procedure yields trajectories in state space, or multidimensional "number lines". We observe that number lines in the data and in the model share three properties: they are *sinuous*, exhibiting oscillations of varying periods that are especially manifest for high numbers; they are *compressed*, with diminishing portions of the curve devoted to increasing numbers; and they are *bounded*, with each trajectory seemingly converging to an attractor state. The two last properties are very much unexpected, but they can be understood within our model in terms of damped oscillations in the von Mises iteration, due to particular eigenvalues of the exponentiated matrix (Quarteroni et al., 2007).

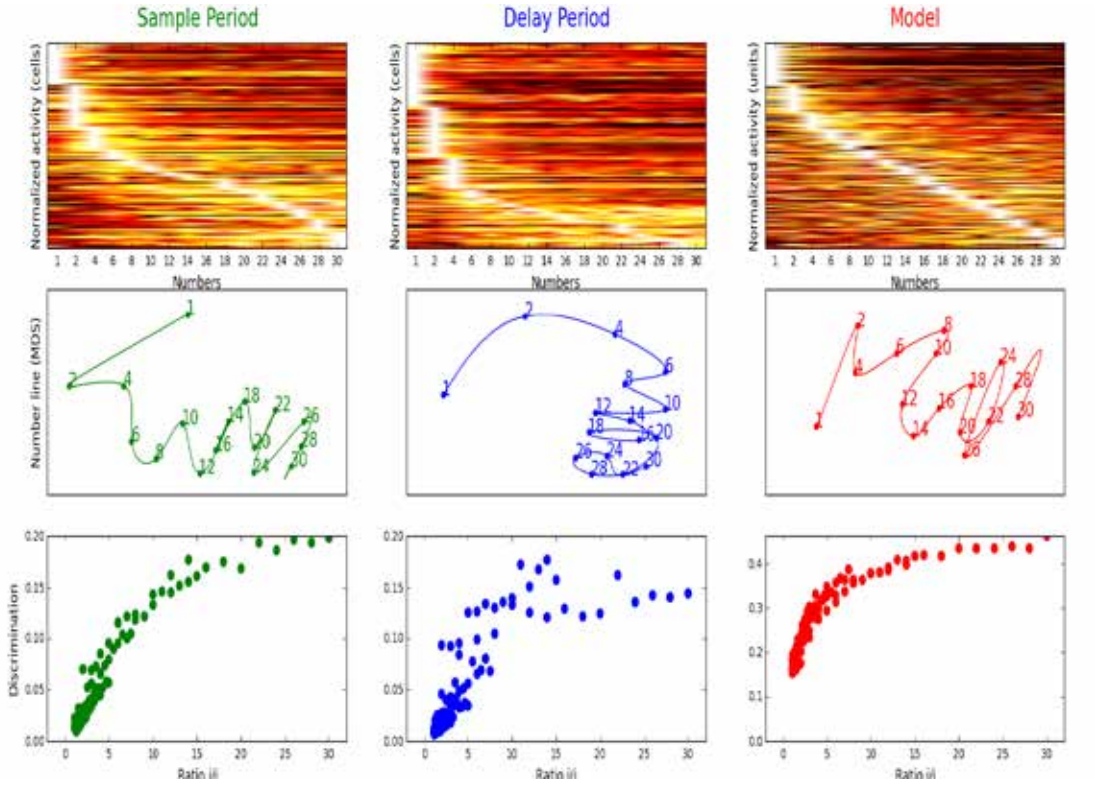


Figure 111

Scaling up to 30 numbers: understanding sequences of number states in the data (columns 1 and 2: analyses based on the normalized firing rates recorded by Nieder and Merten during the sample and delay periods, respectively) and in the model (column 3: analyses based on normalized activations). Line 1: Normalized firing rates for all tested numbers of all PFC cells or model units, ordered by decreasing preferred number. The color scheme maps the rescaled values to a gradient of black (zero), red, yellow and white (one) colors. Line 2: Multidimensional scaling of number states. When interpolated by parameterized polynomials, sequences of number states describe trajectories that are curvilinear, compressed and converging. Line 3: Discrimination (1 - cosine similarity) between number states as a function of their number ratio. The curves for sample data and for the model exhibit a similar shape, consistent with the Weber-Fechner law which holds that discrimination between number states should increase monotonically with the ratio of the numbers being compared.

Finally, we assessed the Weber-Fechner law in the data and in the model. One way to state this law is that the discrimination between two states should increase monotonically with the ratio of the numbers they represent. Figure 111 (bottom line) shows how discrimination changes as a function of number ratio, for the 120 possible pairs of numbers used by (Nieder and Merten, 2007). Although discriminations between number states in the delay period are noisier and harder to interpret, discriminations in the sample period are strikingly similar to those computed based on the model's activations, exhibiting a smooth, monotonic increase as a function of number ratio characteristic of the Weber-Fechner law.

These results show that it is possible to revisit the theory of the number sense in a completely different way. The alternative model can explain the Weber-Fechner law from first principles (i.e. from the linear algebra of iterative matrix powers), and predicts a specific distribution of neural tuning curves which is borne out by electrophysiology.

General summary

We now summarize the output of our work.

A **Dataset Information Card** has been completed (see DIC Task T3.6.1 "Cortical simulations for symbolic and non-symbolic stimuli").

Completeness of data sets and models

No experimental data were collected for this subtask. Our contribution to the study of symbols in the brain was three-fold. First, we have produced a thorough synthesis of the literature on symbol areas in the brain, making specific predictions that are now being tested experimentally, and laying down the requirements for future computational models. Second, we have explored the "shape hypothesis" for the apparition of symbol areas, through a deep convolutional network model of the primate visual system, and using explicit mappings between activations in the network and in the cortex. This work does not present a complete computational account of the emergence of symbol form areas in ventral occipitotemporal cortex, but it provides evidence for the shape hypothesis and clarifies the cortical mechanisms involved. Third, we have devised a computational model based on random matrix theory, which sheds new light on the foundations of the number sense at the neural level.

Location of data sets and models

The OverFeat model is available from the machine learning library "Scikit-Learn Theano", which is hosted and curated by the Parietal research team at Neurospin (Bertrand Thirion, SP2). The network-to-cortical mapping on which our analysis builds is not yet published, but will be deposited at the same location after publication. A dataset containing the cortical simulations from OverFeat for different geometric transforms of faces, tools, houses, letters, strings of letters and strings of pseudoletters, has been deposited at the following address: http://s3.data.kit.edu/3.6.1

Dataset quality, value and usage by others

The predicted cortical activations deposited at the above address constitute a novel effort to make completely explicit the link between computational models (here, a deep convolutional network) and the observables of the brain (voxel activations as revealed by fMRI). This dataset has not been used by other researchers – the corresponding paper is not yet published, being in second revision.

5.2 Linguistic and Non-Linguistic Nested Structures

Task T3.6.2 - Stanislas Dehaene (CEA), Christophe Pallier (CEA), Florent Meyniel (CEA)

Review of the cognitive architectures for the representation of sequences

Stanislas Dehaene, Florent Meyniel, Catherine Wacongne, Liping Wang, Christophe Pallier "The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees", Neuron, Volume 88, Issue 1, p2–19, 7 October 2015

Abstract

A sequence of images, sounds, or words can be stored at several levels of detail, from specific items and their timing to abstract structure. We propose a taxonomy of five distinct cerebral mechanisms for sequence coding: transitions and timing knowledge, chunking, ordinal knowledge, algebraic patterns, and nested tree structures. In each case, we review the available experimental paradigms and list the behavioral and neural signatures of the systems involved. Tree structures require a specific recursive neural code, as yet unidentified by electrophysiology, possibly unique to humans, and which may explain the singularity of human language and cognition.

Data set 1: Encoding of syntactic structures

Sentences are not mere linear strings of words. As detailed in the preceding review, they possess an internal hierarchical structure that expresses the relationships between words (e.g., compare ((black taxi) driver) and (black (taxi driver))). The basic operation that creates the subtrees from these syntactic structures is called “MERGE” in linguistic theory. In a previous (Pallier et al., 2011), we identified a set of brain regions involved in this operation.

Another basic syntactic operation is “MOVE”, whose cerebral bases we explore in the present experiment. According to linguistic theory, MOVE can create, inside a syntactic tree, empty positions marking the location of displaced word See (Figure 112). For example, the french sentence "Je le vois" (meaning 'I see it'; literally 'I it see') results from the movement of the object "le" to a preverbal position, leaving an empty position after the verb. Despite consisting of three words, this sentence needs to have four syntactic positions at a more abstract level. Similarly, in a sentence like "Qui vois-tu?" ("Who do you see?"), it can be argued that the words "Qui" and "vois" have moved to the front of sentence, leaving two empty positions.

Are there brain areas whose activation is driven by the number of abstract syntactic positions ? Do the different types of syntactic movements yield similar or different activations patterns ? To address these questions, we generated sentences belonging to a large variety (35) of syntactic constructions, obtained by combining four types of syntactic movements (Wh-movement, V-movement, Clitic movement, and NP-movement) in different ways (see Table 2). A total of 525 french sentences, 2 to 4 words long were created, which contained 0 to 3 additional empty syntactic positions.

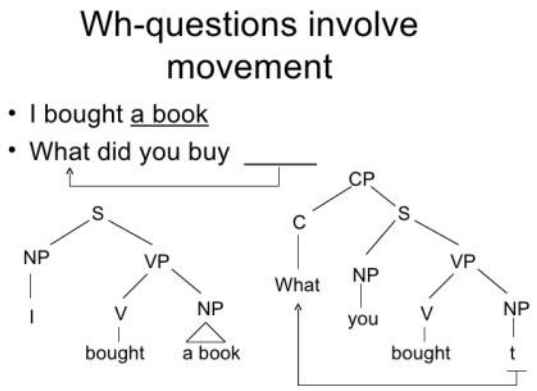


Figure 112: To generate a question, the “MOVE” operation displaces the word “what” from the postverbal position to the beginning of the sentence

1 Argument Unacc	Example
c01_Unacc_Decl_NP_1Ag	Il maigrit.
c02_Unacc_ynQ_NP_1Ag	Tu déchantes
c03_Unacc_Qinv_NP_V_1Ag	Brille-t-elle?

2 Arguments loc	Example
c07_Unacc_Decl_NP_2Ag_loc	Il hiberne là.
c08_Unacc_ynQ_NP_2Ag_loc	Tu vis là?
c09_Unacc_Qinv_NP_V_Wh_2Ag_loc	Où réside-t-elle?
c10_Unacc_Qinv_NP_V_2Ag_loc	Siège-t-il là?
c11_Unacc_Decl_NP_Cl1_2Ag_loc	Tu y survis.
c12_Unacc_ynQ_NP_Cl1_2Ag_loc	Elle y figure?
c13_Unacc_Qinv_NP_V_Cl1_2Ag_loc	Y déteint-t-il?
c14_Unacc_Q_NP_Wh_2Ag_loc	Où elle revient?

3 Arguments obj loc	Example
c23_transy_Decl_3Ag_loc	Il abandonne ça là.
c24_transy_ynQ_3Ag_loc	On adosse ça ici?
c25_transy_Qinv_V_3Ag_loc	Assigne-t-on ça ici?
c26_transy_WhQinv_V_Wh_3Ag_loc	Où colles-tu ça?
c27_transy_Decl_Cl1ob_3Ag_loc	Tu l'enlèves là.
c28_transy_ynQ_Cl1ob_3Ag_loc	On l'adresse là?
c29_transy_Qinv_V_Cl1ob_3Ag_loc	Le retire-t-elle là?
c30_transy_WhQ_Wh_3Ag_loc	Où elle envoie ça?

1 Argument Unerg	Example
c04_Unerg_Decl_1Ag	Tu bailles.
c05_Unerg_ynQ_1Ag	Tu défiles?
c06_Unerg_Qinv_V_1Ag	Dort-elle?

2 Arguments obj	Example
c15_Trans_Decl_2Ag_obj	Tu détruis ça.
c16_Trans_ynQ_2Ag_obj	Il méprise ça?
c17_Trans_Qinv_V_2Ag_obj	Critique-t-il ça?
c18_Trans_whQinv_V_Wh_2Ag_obj	Qui méprise-t-il?
c19_Trans_Decl_Cl1_2Ag_obj	Tu l'esquives.
c20_Trans_ynQ_Cl1_2Ag_obj	Elle l'imite?
c21_Trans_Qinv_V_Cl1_2Ag_obj	L'adopte-t-il?
c22_Trans_whQ_Wh_2Ag_obj	Qui elle écoute?

3 Arguments obj dat clitic	Example
c31_transy_WhQ_Wh_Cl_3Ag_dat	Qui on lui présente?
c32_transy_WhQ_Wh_Cl_V_3Ag_dat	Qui lui soumet-elle?

3_Arguments_2Clitics obj loc	Example
c33_Decl_2ClobjIoc_3Ag_4W_2EC	On l'y joint.
c34_ynQ_2ClobjIoc_3Ag_4W_2EC	Tu l'y appliques?
c35_Qinv_V_2ClobjIoc_3Ag_4w_3EC	L'y associes-tu?

Table 2: The 35 Experimental Conditions Obtained by Combining Several Types of Syntactic Movements.

These sentences were presented visually in a randomized order, to volunteers who simply had to read them while being scanned at 3Telsa. Each sentence was flashed for 200ms with an inter-stimuli interval of 4 seconds. Infrequently, a request to press a response button was displayed in order to make sure that the participants read the stimuli.

Twenty-two native French speakers, all of them right-handed, participated in the experiment which was approved by the regional ethic committee (Comité de Protection des Personnes Ile-de-France VII, Protocole de Recherche Biomedicale #2008-A00241-54/1). Two participants were excluded because of movements of too large amplitude (larger than 1.5mm in translation or 1.5 degree of rotation).

The scanning session started with the acquisition of an anatomical T1-weighted scan (1mm isotropic) for 8min, and continued with five functional MRI sessions of 7 minutes each. An echo planar multiband acquisition sequence allowed us to record the whole brain (80 slices of 1.5mm) every 1.5s with voxels of 1.5mm.

A linear model with the 35 conditions modelled as independent regressors was created in order to estimate the responses to each of them.

In a first, unsupervised, data analysis, we examined the response profiles across the 35 conditions of a set of 14 regions of interest for language processing. A clustering algorithm ran on the matrix of correlations between ROI response profiles, highlighted regions with similar response profiles (see Figure 113). In collaboration with a linguist, Luigi Rizzi, we are currently analysing the response profiles in details to characterise the linguistic factors that explain the behaviour of the regions. Encompassing a wide range of syntactic constructions, these data provide a test bed to test competing linguistic theories.

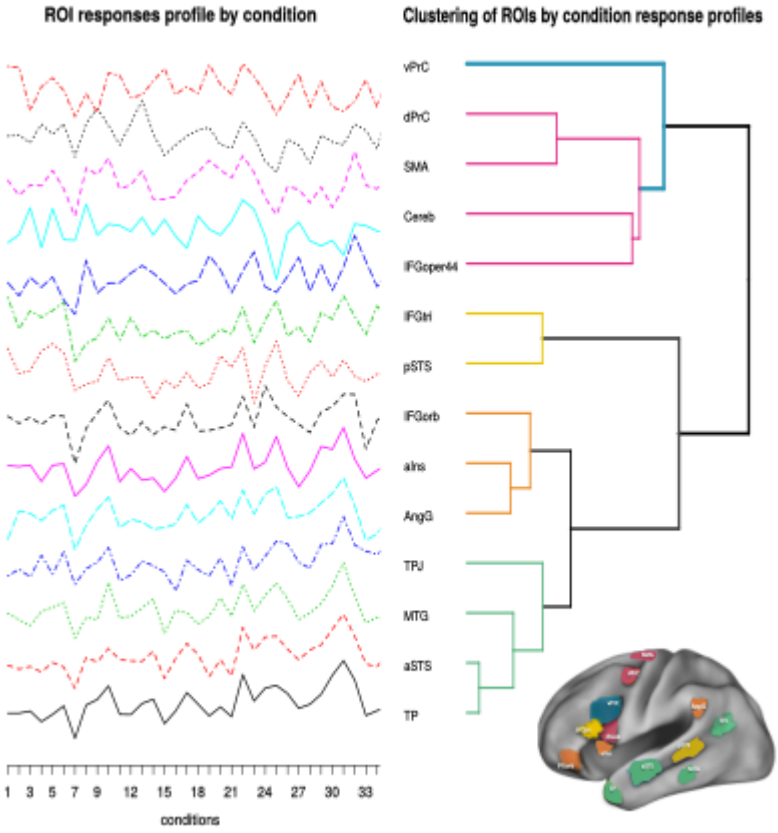


Figure 113: Profiles of responses across the 35 conditions and result of a clustering algorigthm applied to the matrix of correlations of profiles between regions (regions with similar responses are grouped together).

In another, planned analysis, we searched for regions "encoding" the number of abstract positions (overt words + empty positions). A linear contrast with the number of syntactic positions only revealed a single brain area, located in the dorso-precentral gyrus. On closer inspection, more fine-grained contrasts highlighted specific effects of the four different types of movements (Figure 114) showing that Verb and Wh-movements activated similar frontal regions (IFG and Precentral cortex), while Clitic movement modulated activation mostly in the temporo-parietal region and NP movement involved the medial prefrontal cortex. These results suggest that the MOVE operation is not a unitary one.

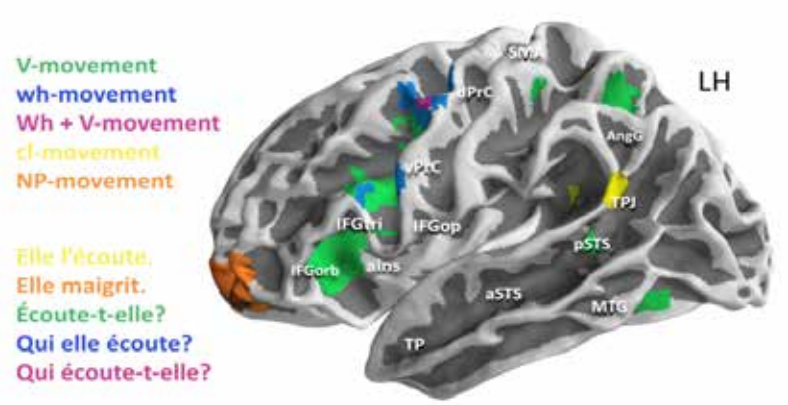


Figure 114: Areas where movements elicited significant activations (overlap of the SPM T maps associated to the 4 different types of movement, thresholded at p<.001 voxelwise uncorrected for multiple comparisons ; at a family-wise corrected level of p<.05, only the V and Wh-movement in the frontal areas remain significant)

A **Dataset Information Card** has been completed (see DIC Task T3.6.2 "Encoding of syntactic structures").

Data Provenance

The data were collected at Neurospin by Christophe PALLIER and Murielle FABRE.

Location of our data storage

The data reside on the acquisition server of the Neurospin center at the CEA. They will be pushed on the HBP server at http://sp3.s3.data.kit.edu/3.6.2

Self-analysis of the value and completeness of our data:

This database is complete, comprising the preprocessed (movement corrected and spatially normalized) scans of 20 participants, as well as the individual first level linear models.

Dissemination:

This study has been presented at the workshop on "new concept in neural pattern encoding" in Gif-sur-Yvette, 28-29 January 2016 (http://news2016.sciencesconf.org/).
The scientific paper is currently in preparation (as of February 2016).

Data set 2: Bayesian Modeling of Expectation Effects in Sequences

Rational of the work

Recent advances in experimental and theoretical neuroscience suggest that the brain is a powerful device that constantly computes the statistics of its environment (Behrens et al., 2007; Dayan and Hinton, 1996; Dehaene et al., 2015; Friston, 2003, 2010; Knill and Pouget, 2004). Experimental evidence relies on signatures of this process that can be found in the brain. Indeed, all learning algorithms learn by estimating whether new observations depart from their current expectations and use this comparison to update their estimates (Rescorla and Wagner, 1972; Sutton and Barto, 1998). The discrepancy between expectation and observation can be formalized with the notion of surprise (O'Reilly et al., 2013; Shannon, 1948; Strange et al., 2005). Several brain signals resemble these theoretical surprise levels in fMRI recordings (Huettel et al., 2002; Pessiglione et al., 2006) or electrophysiological recordings (Kolossa et al., 2013; Lieder et al., 2013; Mars et al., 2008; Squires et al., 1976), so that we may speak of "neural expectation-violation signals". Beside correlative evidence between recorded signals and theoretical measures, pharmacological manipulations provided evidence for a causal role of these neural expectation-violation signals in learning (Aston-Jones and Cohen, 2005; Nieuwenhuis et al., 2005; Pessiglione et al., 2006).

Optimal solutions to reason in uncertain contexts like learning problems are provided by Bayesian models – their optimality derive from their very mathematical foundations (Daunizeau et al., 2010; Jaynes, 2003). We therefore used Bayesian models to compute what can be learned, and hence what can be expected, from a sequential input. Sequences are well suited to study learning from a methodological viewpoint. Thousands of stimuli can easily be presented in typical experimental conditions and in a learning algorithm, each time a new stimulus is received, it is compared with the current expectations. This comparison may elicit neural expectation-violation signals. Numerous neural expectation-violation signals can therefore be recorded and used to "reverse-engineer" the learning algorithm implemented in the brain: by comparing learning algorithms that generate distinct theoretical surprise levels, one can identify the learning algorithm that best correspond to the actual neural expectation-violation signals recorded.

We built on this experimental and conceptual framework to address the following questions:

- What kind of statistics does the brain learn? We tested frequencies vs. transitional probabilities. The difference between frequency and transition probability is that a context is taken into account: the probability of a given stimulus may depend on the preceding one (Strauss et al., 2015; Wacongne et al., 2012). Because of this fundamental distinction, transition probabilities are the first building blocks of the neural representation of sequences (Dehaene et al., 2015).
- What is the temporal horizon of expectation effects? Do expectations build on a short-lived or protracted history of observations? Is this temporal horizon fixed or adapted to the statistics of the input?
- Do several expectation signals co-exist in the brain? At least two neural expectation-violation signals have been identified in electro-physiological recordings. They differ in their latency: an early violation response around 150-250ms, the so-called Mismatch-negativity (MMN) and a late violation response after 300 ms for the P300 or even later (400-600 ms) for the slow-wave (Squires et al., 1976; Strauss et al., 2015; Wacongne et al., 2012). Do they correspond to different expectations?

Experimental and theoretical work

This work was presented in conference posters and the results are in preparation for a publication, we summarize here the main findings. Previous modeling accounts of expectation-violation signals in sequences were based on linear combination of a list of factors. In the seminal work by Squires et al (Squires et al., 1976), this list included the local frequencies of stimuli and their transition (given the recent history), the global frequencies (manipulated experimentally). More recent models include even include additional factors (Kolossa et al., 2013). These factors are essentially descriptive. Bayesian inference on the contrary provides a principled account (Mars et al., 2008) of learning, and hence of expectation-violation signals. We therefore proposed a Bayesian model to tie together the list of factors previously proposed in a very parsimonious manner. This model learns transition probabilities between stimuli, given a recent history of observations. The span of this recent history is the only free parameter of the model. Figure 115 shows that this simple model provides a remarkable fit of the seminal data by Squires et al, when the span history is around 10 stimuli. This model also accounts for other data, such as performance in reaction time tasks, like in the study by Huettel et al (Huettel et al., 2002) in which subjects had to press a left or right button depending of a random sequence of cues.

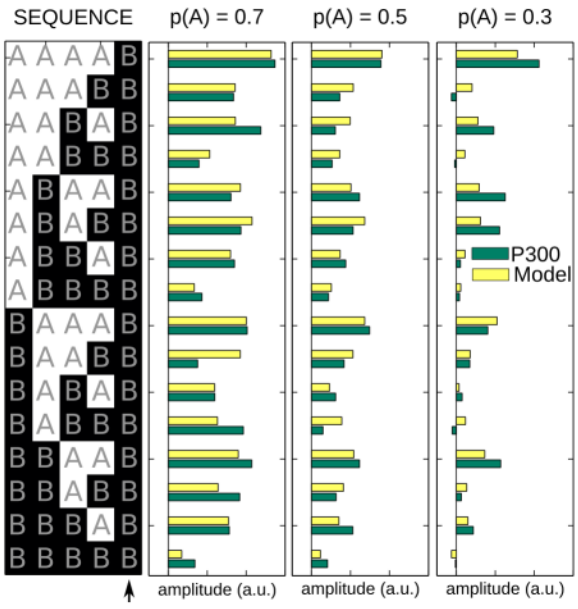


Figure 115: Modelling the P300 as a surprise signal (data from Squires et al., 1976).

Three long sequences of binary stimuli ('A' and 'B') were presented to subjects. These sequences were generated randomly, such that the average proportion of As was a 70%, a 50% or a 30% in distinct sessions. Within each session, the amplitude of the P300 waveform elicited by the B stimuli (green) were averaged after sorting responses based on the preceding pattern of 4 stimuli (the 16 possible patterns are listed line-wise on the left). Average theoretical surprise levels (yellow) were computed similarly for the Bayesian model. Surprise is formally the negative log-likelihood: -log(p), where p is the likelihood of observing B given the past observations. Numeric values were arbitrarily scaled to align the P300 and the model on average. The model reproduces several features of the P300: it is impacted by the global frequency of As (the session type), the local frequency of As within a given pattern and the local pattern of alternation or repetitions.

Because our model learns transition probabilities, it makes predictions that were left untested in the seminal design by Squires et al (1976) and their continuators (Kolossa et al., 2013; Lieder et al., 2013; Mars et al., 2008): responses to the repetition or alternation of stimuli (AA vs. AB) should depend on the overall frequency of AA and AB transitions,

which were not manipulated in previous designs. We therefore modified the original design of Squires et al and generated four sequences based on the following transition probabilities: "frequency bias" with p(A|B) = p(A|A) = 0.3; "no bias" with p(A|B) = p(A|A) = 0.5; "repetition bias" with p(A|B) = p(B|A) = 0.3 and "alternation bias" with p(A|B) = p(B|A) = 0.7. The sequences "no bias" and "frequency bias" correspond to the conditions tested by Squires et al. The two additional types are matched with the "no bias" sequence in terms of stimulus frequencies (a 50% chance of A and B), but their transition probabilities differ.

We recorded 20 subjects with this design in the MEG, and found indeed that the evoked auditory response was modulated by both the frequency of stimuli ("no bias" vs. "frequency bias"), and the transition probabilities between stimuli (see Figure 116). In particular, repetition of a given stimulus elicited more signal than alternation of stimuli when repetitions were rare ("alternation bias"), and it was the opposite when repetitions were frequent ("repetition bias").

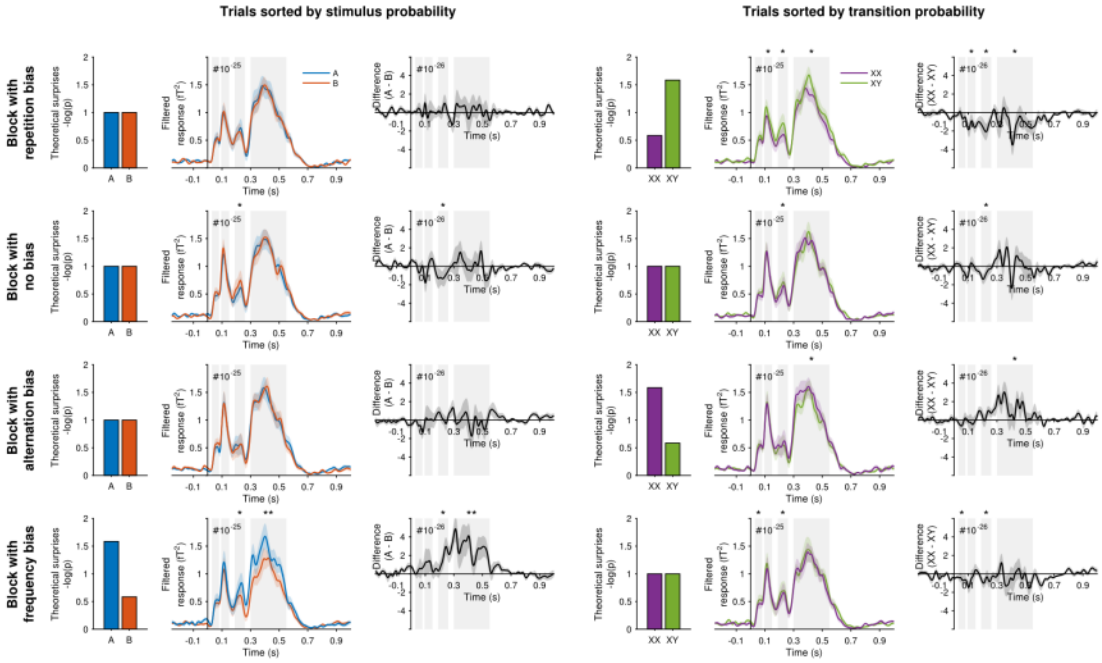


Figure 116: Electrophysiological signatures of expectation-violation in a random sequence of observations.

The figure shows the evoked auditory response sorted by stimuli (left) and transition types (right), in different experimental blocks in which the frequency of stimuli and the transition probabilities between stimuli were manipulated independently. The "evoked auditory response" was computed by filtering, at each time point, the single trial scalp data with the mean topography of the stimulus-locked grand-average over subjects, so that the result is a one-dimension time series. Lines and error shading correspond to mean ± s.e.m.

The results shown in Figure 116 demonstrate an effect of the preceding stimulus on expectation-violation signals. However, our model fit of the data by Squires et al reported above suggests that the brain response to a given stimulus does not only depend on its base rate of occurrence, or which stimulus precedes, but on a longer history of observations. We therefore considered again our Bayesian model that learns transition probabilities and fitted its span for each time point of the evoked-auditory response (see Figure 117A). Our results replicate the well-known distinction between early and late electrophysiological responses in novelty detection. They add to this classic description that the early response reflects a limited integration (around 10 stimuli), and the later response reflects a more protracted integration (around 20-30 stimuli).

After fitting the one-dimensional time series of the evoked-auditory response, we adopted a similar logic to fit the data at each sensor. We also designed additional Bayesian models in order to discriminate between different learning algorithms. One distinction is about the statistics that is learned: instead of learning transition between stimuli, we introduced a variant that learns the frequency of stimuli. Another distinction is about the flexibility of the inference: instead of weighing all stimuli equally within a recent window of observations ("fixed inference"), we introduced a variant that chunks the sequence of observations to maximize locally the likelihood of observations given a specific value of the statistic that is learned ("chunking inference"); this algorithm is detailed in (Meyniel et al., 2015b).

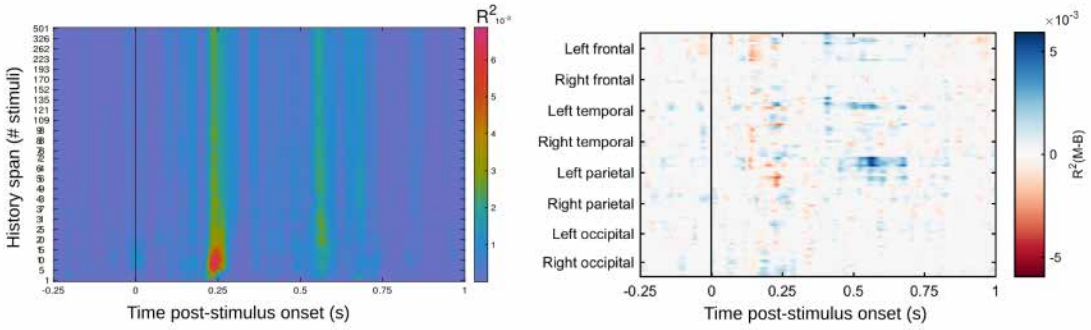


Figure 117: Computational fingerprinting of the novelty detection algorithms.

(A) Fraction of variance explained (R2) in the auditory evoked response, as a function of the history span (the amount of recent stimuli from which statistics are learned) and the latency of the response. The heat map corresponds to the fit of the "repetition bias" and "alternation bias" blocks (similar results were found in the other block types) and the Bayesian model that learns transition probabilities with a "fixed" inference. (B) Comparison between an algorithm that learns transition probabilities (blue) vs. simple frequencies (red), plotted as the difference in fraction of variance explained over the scalp sensors (y-axis) and across time (x-axis). The heat map corresponds to the "repetition bias" and "alternation bias" blocks and the Bayesian model that learns transition probabilities with a "chunking" inference.

Models that learn transition probabilities, rather than frequencies, performed better in general. This was true in particular for sequences in which transition probabilities were biased ("repetition bias" and "alternation bias"), indicating a sensitivity of brain responses to transition probabilities (see Figure 117B). However, the same model also fitted the data recorded in the other sequences slightly better ("frequency bias", "no bias"), indicating that the brain detects local imbalances in transition probabilities, even when this statistic is not biased on the long run. Models that resort to a flexible inference ("chunking") also tended to perform better than those that operate on a fixed history of observation – however further analyses are required at this point.

Research outputs

Here are the main conclusions of our study:

- The brain computes statistics to characterize the sequence of observations that it receives. These statistics are more complex than simple frequencies: they extend to transition probabilities.
- Multiple algorithms operate simultaneously in the brain and transpire into distinct expectation-violation signals. These algorithms differ in the temporal horizon that they consider (recent vs. more protracted history) and the flexibility of this inference (fixed history vs. chunking).

These results were disseminated as follow:

- Preliminary results were presented as a poster (M. Maheu, F. Meyniel and S. Dehaene) at the two-day SP3-SP4 workshop “Probabilistic inference and the brain” (organized by S. Dehaene, F. Meyniel (SP3) and A. Destexhe, W. Maass (SP4 – EITN) at Collège de France, France, September 2015) and at the two-day workshop “New concepts in neural pattern encoding” (Gif-sur-Yvette, France; January 2016)
- Publication of these results was still in preparation (M. Maheu, F. Meyniel and S. Dehaene) in February 2016.

A **Dataset Information Card** has been completed (see DIC Task T3.6.2 “Cortical encoding of probabilistic sequences (MEG and behavior)”).

Data Provenance

The MEG data were collected by Florent MEYNIEL at Neurospin.

Location of our data storage

The data were made available on a server at:

http://s3.data.kit.edu/SP3/3_6_2/Study_ExpectationBayesianModeling_MEG

Data set 3: Encoding of temporal structure by human and non-human primates

Relevant publication: Wang, L., Uhrig, L., Jarraya, B., & Dehaene, S. (2015). Representation of Numerical and Sequential Patterns in Macaque and Human Brains. *Current Biology*, 25(15), 1966-1974.

Our next step was to investigate whether we could identify some level of sequence processing that would be unique to the human brain, and that non-human primates would be unable to attain.

Our logic built upon the local-global test (Bekinschtein et al., 2009), a simple test that we designed to probe sequence representation in human and primate brains. The local-global test consists in exposing subjects to a consistent auditory regularity, and testing the brain's reaction to novel sequences that either respect or violate this regularity. For instance, we first habituate subjects to the repeated hearing of a short sequence of sounds such as *aaaab* (where *a* and *b* are tones or vowels). After habituating to this sequence, we test whether subjects react to the rare presentation of another sequence that violates the template (e.g. *aaaaa*, where the last item is different). Such a violation of the global sequence typically induces a widespread novelty response in temporal, parietal and prefrontal cortices, including pSTS, IPS and IFG. In a recent series of fMRI, ERP and intracranial studies, we have demonstrated this global effect in monkeys,(Uhrig et al., 2014) as well as human adults (Bekinschtein et al., 2009; El Karoui et al., 2014; Strauss et al., 2015; Wacongne et al., 2011) and 3-month-old babies,(Basirat et al., 2014).

This paradigm, however, leads to many additional questions: How is such a sequence encoded? Do monkeys and humans encode it identically? Several possibilities are open: the brain might simply memorize the sequence as a specific melody *aaaab*. Alternatively, it might be sensitive to abstract properties such as number ("four sounds plus another one") or tone-repetition pattern (e.g. "the last sound is different"). The key distinction here is whether an organism uses abstract concepts of number or identity to represent the algebraic pattern underlying a sequence.

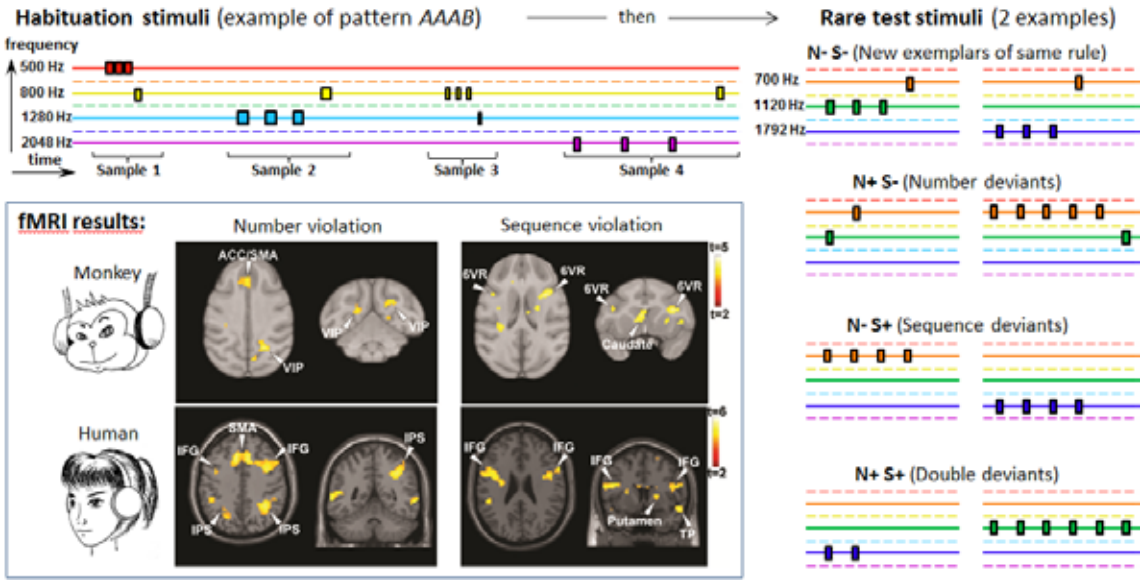


Figure 118: Protocol testing the representation of auditory patterns

We first present several sequences with structure AAAB (top left), then test for fMRI responses to sequences that violate the number of elements, the sound-repetition sequence pattern, or both (right). fMRI results indicate that monkeys and humans do not react to novel exemplars of the same rule (indicating generalizing based on abstract features), but that they react to both types of deviants (indicating a sensitivity to both number and sequence parameters).

To test this idea, we performed an fMRI experiment where monkeys and human adults were tested on a more abstract version of the local-global test, specifically probing the degree of abstractness of their mental representation. They heard a great variety of sequences with a fixed pattern (*AAAB* or *AAAA*). Critically, in this novel design, *A* and *B* could be any of several sounds, and duration and temporal spacing were varied, inciting subjects to memorize the abstract pattern (see Figure 118). Using fMRI, we then tested for brain responses to violations affecting the total number of items (e.g. going from 4 sounds to 2 sounds or to 6 sounds), the sound-repetition pattern (going from *AAAA* to *AAAB* or vice-versa), or both (e.g. going from *AAAA* to *AAAAAB*). Controls ensure that discrimination could not be based on pitch, duration or tempo. Importantly, both species were naïve to the auditory sequences, had not been actively trained to discriminate them, and simply performed an unrelated eye-fixation task while the auditory stimuli were presented.

20 human subjects and 3 monkeys were scanned (please note that since we were specifically prevented from using HBP money for this purpose, the data was acquired using local funds, and has therefore not been made publicly available in the HBP database).

The results indicated the presence of both common and species-specific responses to sequence novelty (See figure 99). First, both monkeys and humans show abstract responses to number (in IPS and ACC/SMA) and to repetition pattern (in basal ganglia, ventral IFG and TP), even when the individual sound change. Thus, even non-human primates are capable of representing the abstract numerical and logical patterns of sequences, as hinted by previous work (Nieder, 2012; Shima et al., 2007). However, the human brain differed from the monkey brain in showing a joint response to both parameters. Tantalizingly, this response occurs only in bilateral IFG (particularly BA 44) and pSTS, and we could prove using single-subject analyses that those sites were also activated by language and music processing. Furthermore, the lack of overlapping responses to numerical and sequence novelty in macaque monkeys was not just a negative result: using representational similarity analysis, we showed that, in the inferior frontal cortex, humans activated positively correlated and therefore similar areas for both types of novelties, while monkeys activated negatively correlated and therefore dissimilar areas, yielding a significant difference between these two species (Figure 119 right).

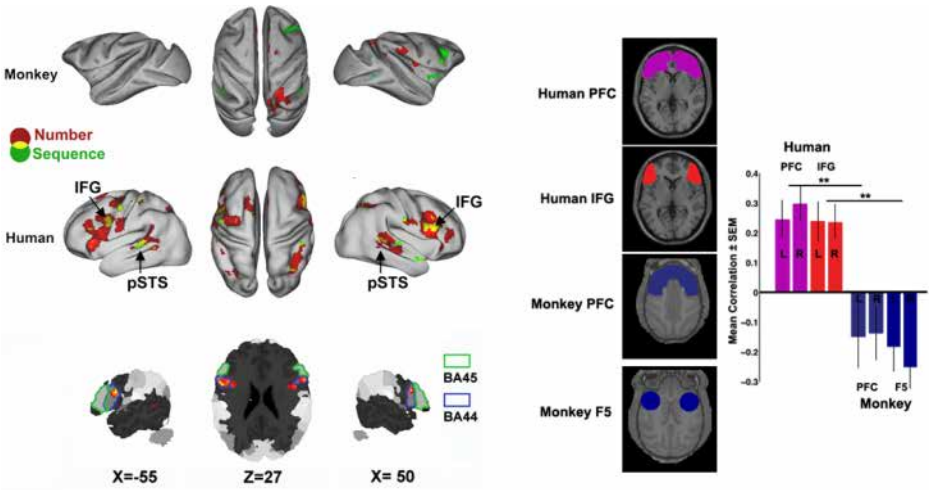


Figure 119: Human-specific integrative representation of auditory patterns

Using the paradigm in Figure 118, fMRI shows that both monkey and human brains react to changes in number (red) and sequence patterns (green). However, only the human brain shows an intersection of both responses (yellow). This intersection occurs in the inferior frontal gyrus (IFG) and posterior STS, at sites overlapping with those engaged in language syntax. Representational similarity analysis (right) confirms that violations of number and sequence co-activate the same voxels in human IFG, suggesting an integrative representation of the pattern underlying the sequences (e.g. “3 tones, then a different one”). Monkeys show non-overlapping

activations to number and to sequence changes, suggesting a sensitivity to isolated features, but no capacity to integrate them into a single unified pattern.

Conclusion. The results, which were published in the journal *Current Biology*, indicate that humans recruit a species-specific circuit, overlapping with language areas, and capable of forming a representation of the whole pattern underlying a sequence. Monkeys encode individual features of the sequence, including abstract ones (e.g. total of 4 sounds; the last sound is different; etc), but if our result is correct, they do not seem to encode the global algebraic pattern ("3 identical sounds followed by a different one").

We have applied to the ERC to continue this research. In the future, we will vary the **complexity** of the learned sequences, by designing **a hierarchy of sequences** dissociating the 5 levels of sequence representation dissected in our review (see above). Subjects will be exposed to short blocks during which most sequences follow some regularity. Rare novel stimuli will allow us to measure generalization and novelty responses. We will monitor the amount, localization and dynamics of brain activation elicited by habituation stimuli, as well as the brain response to deviants. Our pilots and prior work (Bor et al., 2003) indicate that fMRI activation varying with rule complexity can be observed in human and monkey.

Data Provenance. The data were acquired at Neurospin by Liping WANG and Marie AMALRIC.

Dissemination. This work has been published and presented at several international meetings and lab presentations (Tokyo, March 2015; MIT, July 2015; Shanghai, September 2015; etc).

5.3 The Social Brain – Representing the Self in Relation to Others

Task T3.6.3 - Riitta Hari (AALTO), Lauri Parkkonen (AALTO), Linda Henriksson (AALTO)

Review of the cognitive architecture for social representation and social interactions

We have reviewed the cognitive architecture supporting social interaction, and social functions in general, in our recent publication (Hari et al., 2015).

Riitta Hari, Linda Henriksson, Sanna Malinen, Lauri Parkkonen « Centrality of Social Interaction in Human Brain Function", Neuron, Volume 88, Issue 1, p181–193, 7 October 2015

Abstract

People are embedded in social interaction that shapes their brains throughout lifetime. Instead of emerging from lower-level cognitive functions, social interaction could be the default mode via which humans communicate with their environment. Should this hypothesis be true, it would have profound implications on how we think about brain functions and how we dissect and simulate them. We suggest that the research on the brain basis of social cognition and interaction should move from passive spectator science to studies including engaged participants and simultaneous recordings from the brains of the interacting persons.

Review on the importance of timing in brain function

In our other review paper, we have emphasized the importance of timing for brain functions, including those supporting social interaction. Our point is that the structural architecture of the brain (the connectome) needs to be complemented with the dynamics of the connections to understand human brain and behavior (Hari and Parkkonen, 2015).

Data set

Introduction

This task aimed at providing constraints for the emergence and shaping of the conscious mind, specifically trying to argue for the importance of other people in this process. The key scientific problem is whether smooth social interaction is the default mode of human brain function that enables social cognition (as we assume) or whether it is the result of bottom-up computations based on complex cognitive skills.

We designed and conducted neuroimaging experiments on humans to address these constraints. The experimental work included 1) an fMRI localizer for social brain functions using simplified social stimuli, 2) a dual-MEG set-up and measurements to study the brain function of two subjects engaged in real social interaction.

During the HBP Ramp-Up Phase, we have collaborated with CEA (Stanislas Dehaene) for a subset of the stimuli for the social localizer and with EKUT (Martin Giese) for planning action-perception experiments for our dual-MEG set-up.

Design and Results

Publications of dual-MEG results

The construction and application our unique dual-MEG system for simultaneous measurements of two interacting subjects have been described in four original-research publications as of writing this. In the first one (Zhdanov et al., 2015), we describe the dual-MEG set-up and present a proof-of-concept result showing interbrain coherence during synchronous joint hand movements of the two participants.

With the dual-MEG set-up, we have conducted and analysed recordings to study brain mechanisms supporting both nonverbal and verbal interaction (Himberg et al., 2015; Mandel et al., 2016; Zhou et al., 2016).

Experimental design and data set: Social localizer

We have built a comprehensive localizer fMRI experiment for identifying brain regions involved in social cognition. The localizer includes following categories of visual stimuli, with appropriate control conditions (not listed here): biological movement, goal-directed action, tasks related to self vs. other, people in social interaction, theory-of-mind ability (moving geometrical shapes, reading the mind in the eyes), joint attention, faces, and bodies.

We have acquired data from 18 subjects using this localizer; the data acquisition has been funded jointly by HBP and Aalto Brain Centre. By the end of the Ramp-Up Phase of HBP, we deliver (1) the experimental design with the stimuli (copyright restrictions may apply to some of the stimuli) and (2) the key results on brain regions involved in different social tasks. The pre-processed raw fMRI data can also be delivered on request. The current approval from the Ethics committee of Aalto University permits sharing the data with HBP partners but not of other parties. The data are stored at the Aalto University.

Figure 120 shows preliminary results from the social-localizer experiment. Data from 16 subjects were included in the analysis as two subjects had to be excluded due to excessive head motion. All stimuli were shown within one 3T fMRI session and the presentation order of the stimuli was different for each subject. The results are visualized on the inflated cortical surface of the Freesurfer average brain (right hemisphere).

The main analysis of the social-localizer data has been completed and a more detailed

analysis is in progress. The project will be described in the Master's thesis of Timo Nurmi, and a scientific publication will be written about the results during year 2016.

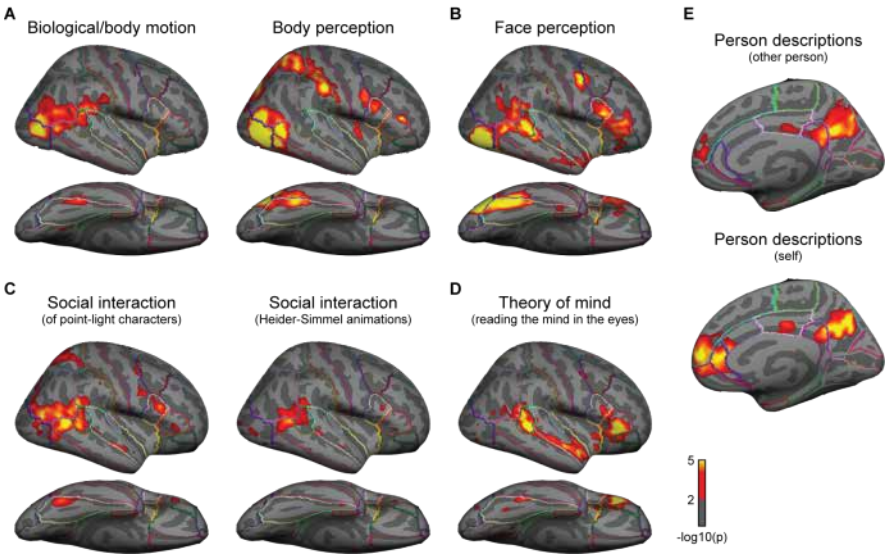


Figure 120: Results from the social-localizer fMRI experiment.

A) The group-level fMRI results for the contrast between biological motion (point-light displays of human movement) and non-biological motion (point-light displays of moving shapes), and for the contrast between pictures of body parts and pictures of objects. B) Contrast between video clips of human faces and natural scenes. C) (left) Contrast between point-light displays of human characters and social (e.g., communicative gestures) and non-social (e.g., walking) situations, and (right) between animations where geometrical shapes were in social interaction (Heider-Simmel animations) and animations of the same geometrical shapes in physical motion. D) Contrast between situations when the subjects were evaluating either the mental or the physical characteristics of a person of whom they saw only the upper face (reading the mind in the eyes test). E) In different task-blocks, the subject judged how well some adjectives described themselves, another person (the president of Finland), or an object (a car). Contrasts are shown between the judgements of (top) other person and the object, and (bottom) the self and the object. All activations are thresholded at $p < 0.01$ (uncorrected); the color lines indicate borders of neuroanatomical regions from an atlas.

Experimental set-up: MEG/EEG imaging of two-person interaction

We have constructed a unique hyperscanning setup, which allows recording MEG/EEG simultaneously from two interacting subjects; see Figure 121. This setup allows connecting two MEG systems at different geographical locations by providing a low-latency audio-visual link and accurate synchronization of the recorded MEG, audio and video data. We have published the setup and proof-of-concept recordings (Zhdanov et al., 2015).

As a Deliverable, we offer the dual-MEG setup to interested HBP partners for joint experimentation. Due to a restrictive ethics permit, the existing dual-MEG recordings unfortunately cannot be shared with groups outside of Aalto University, but future recordings under a new ethics permit could be shared among the HBP partners.

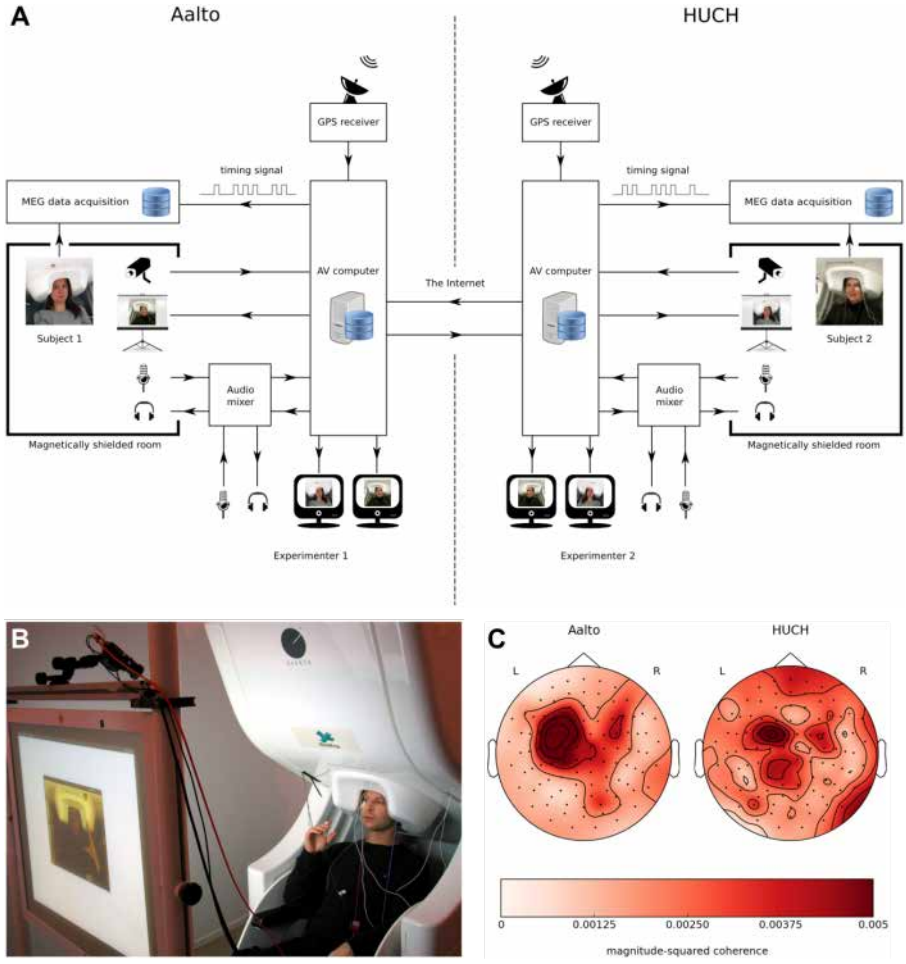


Figure 121

A) Our dual-MEG set-up. The two MEG systems in separate laboratories are connected by an audio-visual link on the Internet, and the data are accurately time-stamped at both laboratories to allow for off-line synchronization of the recordings for joint analysis. B) The subject sees and hears the other subject with an end-to-end delay of about 120 ms. C) Interbrain coherence of MEG signals during mutually synchronized movements of the right hand.

Conclusions

We have designed and tested an fMRI localizer to highlight brain areas specifically relevant for social interaction and mutual understanding. A complete dataset comprising 18 subjects has been acquired. The experimental design and the corresponding dataset will be made available to the HBP consortium; fully open distribution of the fMRI data is not permitted by the Finnish law.

We have also constructed a dual-MEG set-up that allows measuring two interaction subjects at the same time. This set-up is available for interested HBP partners for collaborative experimentation.

Our research supported by HBP has been disseminated as two review papers and four original-research papers, all in peer-reviewed international journals. The data and measurement set-ups have not yet been used by other partners.

A **Dataset Information Card** has been completed (see DIC Task T3.6.3 “fMRI localizer of social brain”).

Data Provenance

Data were collected using a 3-T whole-body MRI scanner (Magnetom Skyra, Siemens) at the Advanced Magnetic Imaging Centre (http://ani.aalto.fi/en/ami_centre/) at *Aalto-korkeakoulusäätiö*, in Finland. The pilot fMRI data were collected by Timo NURMI and Linda HENRIKSSON between October 2014 and March 2015 at the Department of Neuroscience and Biomedical Engineering, *Aalto-korkeakoulusäätiö*.

Annex A: References

Abboud, S., Maidenbaum, S., Dehaene, S., and Amedi, A. (2015). A number-form area in the blind. Nat. Commun. *6*, 6026.

Van Albada, S.J., Helias, M., and Diesmann, M. (2015). Scalability of Asynchronous Networks Is Limited by One-to-One Mapping between Effective Connectivity and Correlations. PLOS Comput. Biol. *11*, e1004490.

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields.

Angelucci, A., and Bullier, J. (2003). Reaching beyond the classical receptive field of V1 neurons: Horizontal or feedback axons? J. Physiol. Paris *97*, 141–154.

Arnulfo, G., Narizzano, M., Cardinale, F., Fato, M.M., and Palva, J.M. (2015a). Automatic segmentation of deep intracerebral electrodes in computed tomography scans. BMC Bioinformatics *16*, 1–12.

Arnulfo, G., Hirvonen, J., Nobili, L., Palva, S., and Palva, J.M. (2015b). Phase and amplitude correlations in resting-state activity in human stereotactical EEG recordings. NeuroImage *112*, 114–127.

Arsenault, J.T., Rima, S., Stemmann, H., and Vanduffel, W. (2014). Role of the primate ventral tegmental area in reinforcement and motivation. Curr. Biol. CB *24*, 1347–1353.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. NeuroImage *38*, 95–113.

Ashby, F.G., Prinzmetal, W., Ivry, R., and Maddox, W.T. (1996). A formal theory of feature binding in object perception. Psychol. Rev. *103*, 165–192.

Astafiev, S.V., Snyder, A.Z., Shulman, G.L., and Corbetta, M. (2010). Comment on "Modafinil shifts human locus coeruleus to low-tonic, high-phasic activity during functional MRI" and "Homeostatic sleep pressure and responses to sustained attention in the suprachiasmatic area." Science *328*, 309.

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

De Baene, W., and Vogels, R. (2010). Effects of Adaptation on the Stimulus Selectivity of Macaque Inferior Temporal Spiking Activity and Local Field Potentials. Cereb. Cortex *20*, 2145–2165.

Baldauf, D., and Desimone, R. (2014). Neural Mechanisms of Object-Based Attention. Science *25*, 1080–1092.

Banakou, D., and Slater, M. (2014). Body ownership causes illusory self-attribution of speaking and influences subsequent real speaking. Proc. Natl. Acad. Sci. *111*, 17678–17683.

Baraduc, P., Thobois, S., Gan, J., Broussolle, E., and Desmurget, M. (2013). A Common Optimization Principle for Motor Execution in Healthy Subjects and Parkinsonian Patients. J. Neurosci. *33*, 665–677.

Baranski, J.V., and Petrusic, W.M. (1994). The calibration and resolution of confidence in perceptual judgments. Percept. Psychophys. *55*, 412–428.

Barbieri, A., Mazzoni, A., Logothetis, N.K., Panzeri, S., and Brunel, N. (2013). Input dependence of local field potential spectra: experiment vs theory. BMC Neurosci. *14*, 38–39.

Bartos, M., Vida, I., and Jonas, P. (2007). Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. Nat. Rev. Neurosci. *8*, 45–56.

Basirat, A., Dehaene, S., and Dehaene-Lambertz, G. (2014). A hierarchy of cortical responses to sequence violations in three-month-old infants. Cognition *132*, 137–150.

Bastos, A.M., Briggs, F., Alitto, H.J., Mangun, G.R., and Usrey, W.M. (2014). Simultaneous recordings from the primary visual cortex and lateral geniculate nucleus reveal rhythmic interactions and a cortical source for gamma-band oscillations. J. Neurosci. Off. J. Soc. Neurosci. *34*, 7639–7644.

Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J.R., De Weerd, P., Kennedy, H., and Fries, P. (2015a). Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. Neuron *85*, 390–401.

Bastos, A.M., Litvak, V., Moran, R., Bosman, C.A., Fries, P., and Friston, K.J. (2015b). A DCM study of spectral asymmetries in feedforward and feedback connections between visual areas V1 and V4 in the monkey. Neuroimage *108*, 460–475.

Behrens, T.E.J., Woolrich, M.W., Walton, M.E., and Rushworth, M.F.S. (2007). Learning the value of information in an uncertain world. Nat. Neurosci. *10*, 1214–1221.

Behzadi, Y., Restom, K., Liau, J., and Liu, T.T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. NeuroImage *37*, 90–101.

Bekinschtein, T.A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. Proc Natl Acad Sci USA *106*, 1672–1677.

Bendels, M.H.K., and Leibold, C. (2007). Generation of theta oscillations by weakly coupled neural oscillators in the presence of noise. J. Comput. Neurosci. *22*, 173–189.

Bergström, F., and Eriksson, J. (2014). Maintenance of non-consciously presented information engages the prefrontal cortex. Front. Hum. Neurosci. *8*, 938: 1–10.

Berridge, K.C. (2004). Motivation concepts in behavioral neuroscience. Physiol. Behav. *81*, 179–209.

Berridge, K.C. (2007). The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology (Berl.) *191*, 391–431.

Binzegger, T., Douglas, R.J., and Martin, K. a C. (2004). A quantitative map of the circuit of cat primary visual cortex. J. Neurosci. Off. J. Soc. Neurosci. *24*, 8441–8453.

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. Nat. Rev. Neurosci. *13*, 556–571.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J.D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychol. Rev. *113*, 700.

Bollen, K. (2002). Latent Variables in Psychology and the Social Sciences. Annu. Rev. Psychol. *53*, 605–634.

Bor, D., Duncan, J., Wiseman, R.J., and Owen, A.M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. Neuron *37*, 361–367.

Börgers, C., and Kopell, N. (2005). Effects of noisy drive on rhythms in networks of excitatory and inhibitory neurons. Neural Comput. *17*, 557–608.

Bosman, C.A., Schoffelen, J.-M., Brunet, N., Oostenveld, R., Bastos, A.M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., and Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. Neuron *75*, 875–888.

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.

Breakspear, M., Heitmann, S., and Daffertshofer, A. (2010). Generative models of cortical oscillations: neurobiological implications of the kuramoto model. Front Hum Neurosci *4*, 190.

Brennan, A. (2012). Mirror Writing and Hand Dominance in Children: A New Perspective on Motor and Perceptual Theories. Yale Rev. Undergrad. Res. Psychol. 12–23.

Bressler, S.L., and Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn. Sci. *14*, 277–290.

Brown, R.G., and Pluck, G. (2000). Negative symptoms: the “pathology” of motivation and goal-directed behaviour. Trends Neurosci. *23*, 412–417.

Brown, C.A., Campbell, M.C., Karimi, M., Tabbal, S.D., Loftin, S.K., Tian, L.L., Moerlein, S.M., and Perlmutter, J.S. (2012). Dopamine pathway loss in nucleus accumbens and ventral tegmental area predicts apathetic behavior in MPTP-lesioned monkeys. Exp. Neurol. *236*, 190–197.

Brunel, N. (2000). Dynamics of sparsely connected networls of excitatory and inhibitory neurons. Comput. Neurosci. *8*, 183–208.

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*, 415–430.

Brunton, B.W., Botvinick, M.M., and Brody, C.D. (2013). Rats and humans can optimally accumulate evidence for decision-making. Science *340*, 95–98.

Buffalo, E.A., Fries, P., Landman, R., Buschman, T.J., and Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. Proc. Natl. Acad. Sci. U. S. A. *108*, 11262–11267.

Buia, C., and Tiesinga, P. (2006). Attentional modulation of firing rate and synchrony in a model cortical network. J. Comput. Neurosci. *20*, 247–264.

Burgess, N., and O'Keefe, J. (1996). Neuronal computations underlying the firing of place cells and their role in navigation. Hippocampus *6*, 749–762.

Burns, S.P., Xing, D., and Shapley, R.M. (2011). Is gamma-band activity in the local field potential of V1 cortex a "clock" or filtered noise? J. Neurosci. Off. J. Soc. Neurosci. *31*, 9658–9664.

Busemeyer, J.R., and Townsend, J.T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychol. Rev. *100*, 432.

Buzsáki, G. (2006). Rhythms of the Brain.

Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. Science *304*, 1926–1929.

Buzsáki, G., and Wang, X.-J. (2012). Mechanisms of Gamma Oscillations. Annu. Rev. Neurosci. *35*, 203–225.

Buzsáki, G., Geisler, C., Henze, D.A., and Wang, X.-J. (2004). Interneuron Diversity series: Circuit complexity and axon wiring economy of cortical interneurons. Trends Neurosci. *27*, 186–193.

Buzsáki, G., Anastassiou, C. a, and Koch, C. (2012a). The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. *13*, 407–420.

Buzsáki, G., Anastassiou, C.A., and Koch, C. (2012b). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. *13*, 407–420.

Caggiano, V., Pomper, J.K., Fleischer, F., Fogassi, L., Giese, M., and Thier, P. (2013). Mirror neurons in monkey area F5 do not adapt to the observation of repeated actions. Nat Commun *4*, 1433.

Cardin, J.A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L.-H., and Moore, C.I. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. Nature *459*, 663–667.

Caspers, S., Geyer, S., Schleicher, A., Mohlberg, H., Amunts, K., and Zilles, K. (2006). The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual variability. NeuroImage *33*, 430–448.

Cavallari, S., Panzeri, S., and Mazzoni, A. (2014). Comparison of the dynamics of neural interactions between current-based and conductance-based integrate-and-fire recurrent networks. Front. Neural Circuits *8*, 12.

Chen, K., and Wang, D. (2002). A dynamically coupled neural oscillator network for image segmentation. Neural Netw. *15*, 423–439.

Chersi, F., and Burgess, N. (2015). The Cognitive Architecture of Spatial Navigation: Hippocampal and Striatal Contributions. Neuron *88*, 64–77.

Christensen, A., Ilg, W., and Giese, M.A. (2011). Spatiotemporal tuning of the facilitation of biological motion perception by concurrent motor execution. J Neurosci *31*, 3493–3499.

Christensen, A., Giese, M.A., Sultan, F., Mueller, O.M., Goericke, S.L., Ilg, W., and Timmann, D. (2014). An intact action-perception coupling depends on the integrity of the cerebellum. J Neurosci *34*, 6707–6716.

Cichy, R.M., Pantazis, D., and Oliva, A. (2014). Resolving human object recognition in space and time. Nat. Neurosci. *17*, 455–462.

Cisek, P., and Kalaska, J.F. (2010). Neural Mechanisms for Interacting with a World Full of Action Choices. Annu. Rev. Neurosci. Vol 33 *33*, 269–298.

Clayton, M.S., Yeung, N., and Cohen Kadosh, R. (2016). The roles of cortical oscillations in sustained attention. Trends Cogn. Sci. *19*, 188–195.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.A., and Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain J. Neurol. *123 (Pt 2)*, 291–307.

Cohen, L., Jobert, A., Le Bihan, D., and Dehaene, S. (2004). Distinct unimodal and multimodal regions for word processing in the left temporal cortex. NeuroImage *23*, 1256–1270.

Coombes, S., and Bressloff, P.C. (1999). Mode locking and Arnold tongues in integrate-and-fire neural oscillators. Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top. *60*, 2086–2096.

Coombes, S., Schmidt, H., and Bojak, I. (2012). Interface dynamics in planar neural field models. J. Math. Neurosci. *2*, 9.

Cottam, J.C.H., Smith, S.L., and Häusser, M. (2013). Target-specific effects of somatostatin-expressing interneurons on neocortical visual processing. J. Neurosci. Off. J. Soc. Neurosci. *33*, 19567–19578.

Czernecki, V., Schüpbach, M., Yaici, S., Lévy, R., Bardinet, E., Yelnik, J., Dubois, B., and Agid, Y. (2008). Apathy following subthalamic stimulation in Parkinson disease: a dopamine responsive symptom. Mov. Disord. Off. J. Mov. Disord. Soc. *23*, 964–969.

Damier, P., Hirsch, E.C., Agid, Y., and Graybiel, A.M. (1999). The substantia nigra of the human brain. II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. Brain J. Neurol. *122 (Pt 8)*, 1437–1448.

D'Ardenne, K., McClure, S.M., Nystrom, L.E., and Cohen, J.D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science *319*, 1264–1267.

Daunizeau, J., den Ouden, H.E.M., Pessiglione, M., Kiebel, S.J., Stephan, K.E., and Friston, K.J. (2010). Observing the Observer (I): Meta-Bayesian Models of Learning and Decision-Making. PLoS ONE *5*, e15554.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Model-based influences on humans' choices and striatal prediction errors. Neuron *69*, 1204–1215.

Dayan, P. (2012). Twenty-five lessons from computational neuromodulation. Neuron *76*, 240–256.

Dayan, P., and Hinton, G.E. (1996). Varieties of Helmholtz Machine. Neural Netw. *9*, 1385–1403.

D. Brewster (1847). On the conversion of relief by inverted vision (Edinburgh Philosophical Transactions).

deCharms, R.C., and Merzenich, M.M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing. Nature *381*, 610–613.

Deco, G., Pérez-Sanagustín, M., de Lafuente, V., and Romo, R. (2007). Perceptual detection as a dynamical bistability phenomenon: a neurocomputational correlate of sensation. Proc. Natl. Acad. Sci. U. S. A. *104*, 20073–20077.

Deco, G., Jirsa, V., McIntosh, a R., Sporns, O., and Kötter, R. (2009). Key role of coupling, delay, and noise in resting brain fluctuations. Proc. Natl. Acad. Sci. U. S. A. *106*, 10302–10307.

Dehaene, S., and Changeux, J.P. (1993). Development of elementary numerical abilities: a neuronal model. J. Cogn. Neurosci. *5*, 390–407.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. Neuron *88*, 2–19.

Demanet, J., Muhle-Karbe, P.S., Lynn, M.T., Blotenberg, I., and Brass, M. (2013). Power to the will: How exerting physical effort boosts the sense of agency. Cognition *129*, 574–578.

Denk, F., Walton, M.E., Jennings, K.A., Sharp, T., Rushworth, M.F.S., and Bannerman, D.M. (2005). Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. Psychopharmacology (Berl.) *179*, 587–596.

Desikan, R.S., S??gonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al. (2006a). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage *31*, 968–980.

Desikan, R.S., S??gonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al. (2006b). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage *31*, 968–980.

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. NeuroImage *53*, 1–15.

DeWitt, I., and Rauschecker, J.P. (2012). Phoneme and word recognition in the auditory ventral stream. Proc. Natl. Acad. Sci. U. S. A. *109*, E505–E514.

Dodge, Y. (2006). The Oxford Dictionary of Statistical Terms (Oxford University Press).

Donner, T.H., and Nieuwenhuis, S. (2013). Brain-wide gain modulation: the rich get richer. Nat. Neurosci. *16*, 989–990.

Donner, T.H., and Siegel, M. (2011). A framework for local cortical oscillation patterns. Trends Cogn. Sci. *15*, 191–199.

Donner, T.H., Siegel, M., Oostenveld, R., Fries, P., Bauer, M., and Engel, A.K. (2007). Population activity in the human dorsal pathway predicts the accuracy of visual motion detection. J. Neurophysiol. *98*, 345–359.

Donner, T.H., Siegel, M., Fries, P., and Engel, A.K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. Curr. Biol. *19*, 1581–1585.

Dotson, N.M., Salazar, R.F., and Gray, C.M. (2014a). Frontoparietal correlation dynamics reveal interplay between integration and segregation during visual working memory. J. Neurosci. Off. J. Soc. Neurosci. *34*, 13600–13613.

Dotson, N.M., Salazar, R.F., and Gray, C.M. (2014b). Frontoparietal correlation dynamics reveal interplay between integration and segregation during visual working memory. J. Neurosci. Off. J. Soc. Neurosci. *34*, 13600–13613.

Dotson, N.M., Goodell, B., Salazar, R.F., Hoffman, S.J., and Gray, C.M. (2015). Methods, caveats and the future of large-scale microelectrode recordings in the non-human primate. Front. Syst. Neurosci. *9*, 149–149.

Douglas, R.J., and Martin, K.A.C. (2004). Neuronal Circuits of the Neocortex. Annu. Rev. Neurosci. *27*, 419–451.

Du, G., Lewis, M.M., Sen, S., Wang, J., Shaffer, M.L., Styner, M., Yang, Q.X., and Huang, X. (2012). Imaging nigral pathology and clinical progression in Parkinson's disease. Mov. Disord. Off. J. Mov. Disord. Soc. *27*, 1636–1643.

Eckert, M.A., Keren, N.I., and Aston-Jones, G. (2010). "Looking forward with the Locus Coeruleus" eLetter to "Comment on 'Modafinil Shifts Human Locus Coeruleus to Low-Tonic, High-Phasic Activity During Functional MRI' and 'Homeostatic Sleep Pressure and Responses to Sustained Attention in the Suprachiasmatic Area.'" Science *328*, 309.

Eckhorn, R. (1999). Neural mechanisms of scene segmentation: recordings from the visual cortex suggest basic circuits for linking field models. IEEE Trans. Neural Netw. Publ. IEEE Neural Netw. Counc. *10*, 464–479.

Ehringer, H., and Hornykiewicz, O. (1960). [Distribution of noradrenaline and dopamine (3-hydroxytyramine) in the human brain and their behavior in diseases of the extrapyramidal system]. Klin. Wochenschr. *38*, 1236–1239.

Eickenberg, M., Gramfort, A., Varoquaux, G., and Thirion, B. (submitted). Seeing it all: Convolutional Neural Nets Map the Function of the Human Visual System.

Eickhoff, S.B., Schleicher, A., Zilles, K., and Amunts, K. (2006). The human parietal operculum. I. Cytoarchitectonic mapping of subdivisions. Cereb. Cortex N. Y. N 1991 *16*, 254–267.

Eickhoff, S.B., Jbabdi, S., Caspers, S., Laird, A.R., Fox, P.T., Zilles, K., and Behrens, T.E.J. (2010). Anatomical and functional connectivity of cytoarchitectonic areas within the human parietal operculum. J. Neurosci. Off. J. Soc. Neurosci. *30*, 6409–6421.

Einevoll, G.T., Kayser, C., Logothetis, N.K., and Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. Nat. Rev. Neurosci. *14*, 770–785.

El Karoui, I., King, J.-R., Sitt, J., Meyniel, F., Van Gaal, S., Hasboun, D., Adam, C., Navarro, V., Baulac, M., Dehaene, S., et al. (2014). Event-Related Potential, Time-frequency, and Functional Connectivity Facets of Local and Global Auditory Novelty Processing: An Intracranial Study in Humans. Cereb. Cortex N. Y. N 1991.

Engel, A.K., Kreiter, A.K., König, P., and Singer, W. (1991). Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. Proc. Natl. Acad. Sci. U. S. A. *88*, 6048–6052.

Engel, a K., Fries, P., König, P., Brecht, M., and Singer, W. (1999). Temporal binding, binocular rivalry, and consciousness. Conscious. Cogn. *8*, 128–151.

Erlhagen, W., and Schoner, G. (2002). Dynamic field theory of movement preparation. Psychol Rev *109*, 545–572.

Ermentrout, G.B., and Kleinfeld, D. (2001). Traveling electrical waves in cortex: insights from phase dynamics and speculation on a computational role. Neuron *29*, 33–44.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., and Ugurbil, K. (2013). The WU-Minn Human Connectome Project: an overview. NeuroImage *80*, 62–79.

Fedorov, L. (2014). Neurodynamical model for the multi-stable perception of biological motion. (Journal of Vision), p. 1007.

Fedorov, L., and Giese, M.A. (2015). Shading cues in the perception of biological motion: a neural model and a new illusion. Eur. Conf. Vis. Percept. ECVP.

Feng, W., Havenith, M.N., Wang, P., Singer, W., and Nikolić, D. (2010). Frequencies of gamma/beta oscillations are stably tuned to stimulus properties. Neuroreport *21*, 680–684.

Fields, R.D. (2008). Oligodendrocytes changing the rules: action potentials in glia and oligodendrocytes controlling action potentials. Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry *14*, 540–543.

Fleischer, F., Christensen, A., Caggiano, V., Thier, P., and Giese, M.A. (2012). Neural theory for the perception of causal actions. Psychol Res *76*, 476–493.

Fleischer, F., Caggiano, V., Thier, P., and Giese, M.A. (2013). Physiologically inspired model for the visual recognition of transitive hand actions. J Neurosci *33*, 6563–6580.

Fried, I., Wilson, C.L., Maidment, N.T., Engel Jr, J., Behnke, E., Fields, T.A., Macdonald, K.A., Morrow, J.W., and Ackerson, L. (1999). Cerebral microdialysis combined with single-neuron and electroencephalographic recording in neurosurgical patients: technical note. J. Neurosurg. *91*, 697–705.

Fries, P. (2005a). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn. Sci. *9*, 474–480.

Fries, P. (2005b). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn. Sci. *9*, 474–480.

Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. Annu. Rev. Neurosci. *32*, 209–224.

Fries, P. (2015a). Rhythms for Cognition: Communication through Coherence. Neuron *88*, 220–235.

Fries, P. (2015b). Rhythms for Cognition: Communication through Coherence. Neuron *88*, 220–235.

Fries, P., Nikolić, D., and Singer, W. (2007). The gamma cycle. Trends Neurosci. *30*, 309–316.

Friston, K. (2003). Learning and inference in the brain. Neural Netw. Off. J. Int. Neural Netw. Soc. *16*, 1325–1352.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. *11*, 127–138.

Gabitov, E., Manor, D., and Karni, A. (2014). Done that: Short- term Repetition Related Modulation of Motor Cortex Activity as a Stable Signature for Overnight Motor Memory Consolidation. J. Cogn. Neurosci. *26*, 2716–2734.

Gabitov, E., Manor, D., and Karni, A. (2015). Patterns of Modulation in the Activity and Connectivity of Motor Cortex during the Repeated Generation of Movement Sequence. J. Cogn. Neurosci. *27*, 736–751.

Gabitov, E., Manor, D., and Karni, A. (2016). Learning from the Other Limb's Experience: Sharing the "Trained" M1's Representation of the Motor Sequence Knowledge. J. Physiol. *1*, 1–39.

Galán, R., Ermentrout, G., and Urban, N. (2005). Efficient Estimation of Phase-Resetting Curves in Real Neurons and its Significance for Neural-Network Modeling. Phys. Rev. Lett. *94*, 158101.

Gan, J.O., Walton, M.E., and Phillips, P.E.M. (2010). Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. Nat. Neurosci. *13*, 25–27.

Gee, J.W. de, Knapen, T., and Donner, T.H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. Proc. Natl. Acad. Sci. U. S. A. *111*, E618–E625.

Ghosh, A., Rho, Y., McIntosh, a R., Kötter, R., and Jirsa, V.K. (2008). Noise during rest enables the exploration of the brain's dynamic repertoire. PLoS Comput. Biol. *4*, e1000196.

Giese, M.A. (2014). Skeleton Model for the Neurodynamics of Visual Action Representations. In Artificial Neural Networks and Machine Learning – ICANN 2014, S. Wermter, C. Weber, W. Duch, T. Honkela, P. Koprinkova-Hristova, S. Magg, G. Palm, and A.E.P. Villa, eds. (Springer International Publishing), pp. 707–714.

Giese, M.A., and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. Nat. Rev. Neurosci. *4*, 179–192.

Giese, M., Fedorov, L., and Vogels, R. (2015a). Interaction between adaptation and perceptual multi-stability in body motion recognition. J. Vis. *15*, 557.

Giese, M.A., Fedorov, L., and Vogels, R. (2015b). Neural model for multi-stability in visual action recognition. (Prague),.

Gieselmann, M. a., and Thiele, a. (2008). Comparison of spatial integration and surround suppression characteristics in spiking activity and the local field potential in macaque V1. Eur. J. Neurosci. *28*, 447–459.

Giovanni Pezzulo, D.O. (2012). Proactive action preparation: seeing action preparation as a continuous and proactive process. Motor Control *16*, 386–424.

Glover, G.H., Li, T.-Q., and Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn. Reson. Med. *44*, 162–167.

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

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

Gonen, T., Admon, R., Podlipsky, I., and Hendler, T. (2012). From animal model to human brain networking: dynamic causal modeling of motivational systems. J. Neurosci. *32*, 7218–7224.

Gonen, T., Soreq, E., Eldar, E., Ben-Simon, E., Raz, G., and Hendler, T. (2016). Human mesostriatal response tracks motivational tendencies under naturalistic goal conflict. Soc. Cogn. Affect. Neurosci.

Gorell, J.M., Ordidge, R.J., Brown, G.G., Deniau, J.C., Buderer, N.M., and Helpern, J.A. (1995). Increased iron-related MRI contrast in the substantia nigra in Parkinson's disease. Neurology *45*, 1138–1143.

Gould, I.C., Nobre, A.C., Wyart, V., and Rushworth, M.F. (2012). Effects of decision variables and intraparietal stimulation on sensorimotor oscillatory activity in the human brain. J. Neurosci. Off. J. Soc. Neurosci. *32*, 13805–13818.

Gray, J.A., and McNaughton, N. (2003). The neuropsychology of anxiety: An enquiry into the function of the septo-hippocampal system (Oxford university press).

Gray, C.M., Koenig, P., Engel, A.K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects. Nature *338*, 334–337.

Gregoriou, G.G., Gotts, S.J., Zhou, H., and Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. Science *324*, 1207–1210.

Grimaldi, P., Lau, H., and Basso, M.A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. Neurosci. Biobehav. Rev. *55*, 88–97.

Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., and Schnitzler, A. (2004). Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. Proc. Natl. Acad. Sci. U. S. A. *101*, 13050–13055.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions. Biol. Cybern. *23*, 187–202.

Guigon, E., Baraduc, P., and Desmurget, M. (2007). Computational Motor Control: Redundancy and Invariance. J. Neurophysiol. *97*, 331–347.

Hadjipapas, A., Casagrande, E., Nevado, A., Barnes, G.R., Green, G., and Holliday, I.E. (2009). Can we observe collective neuronal activity from macroscopic aggregate signals? NeuroImage *44*, 1290–1303.

Hadjipapas, A., Lowet, E., Roberts, M.J., Peter, A., and De Weerd, P. (2015). Parametric variation of gamma frequency and power with luminance contrast: A comparative study of human MEG and monkey LFP and spike responses. NeuroImage *112*, 327–340.

Haegens, S., Nacher, V., Hernandez, A., Luna, R., Jensen, O., and Romo, R. (2011). Beta oscillations in the monkey sensorimotor network reflect somatosensory decision making. Proc. Natl. Acad. Sci. U. S. A. *108*, 10708–10713.

Hagen, E., Dahmen, D., Stavrinou, M.L., Tetzlaff, T., Diesmann, M., and Einevoll, G.T. (2015). Hybrid scheme for modeling local field potentials from point-neuron networks. *c*, 1–52.

Haider, B., Häusser, M., and Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. Nature *493*, 97–100.

Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev. Mod. Phys. *65*, 413–497.

Hannagan, T., Amedi, A., Cohen, L., Dehaene-Lambertz, G., and Dehaene, S. (2015). Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. Trends Cogn. Sci. *19*, 374–382.

Hari, R., and Parkkonen, L. (2015). The brain timewise: how timing shapes and supports brain function. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *370*.

Hari, R., Henriksson, L., Malinen, S., and Parkkonen, L. (2015). Centrality of Social Interaction in Human Brain Function. Neuron *88*, 181–193.

Harris, C.M., and Wolpert, D.M. (1998). Signal-dependent noise determines motor planning. Nature *394*, 780–784.

Harris, K.D., and Thiele, A. (2011). Cortical state and attention. Nat. Rev. Neurosci. *12*, 509–523.

Hartmann, M.N., Hager, O.M., Tobler, P.N., and Kaiser, S. (2013). Parabolic discounting of monetary rewards by physical effort. Behav. Processes *100*, 192–196.

Hebart, M.N., Donner, T.H., and Haynes, J.-D. (2012). Human visual and parietal cortex encode visual choices independent of motor plans. NeuroImage *63*, 1393–1403.

Hebart, M.N., Schriever, Y., Donner, T.H., and Haynes, J.-D. (2014). The Relationship between Perceptual Decision Variables and Confidence in the Human Brain. Cereb. Cortex.

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

Himberg, T., Hirvenkari, L., Mandel, A., and Hari, R. (2015). Word-by-word entrainment of speech rhythm during joint story building. Name Front. Psychol. *6*, 797.

Hipp, J.F., Engel, A.K., and Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. Neuron *69*, 387–396.

Hirsch, E., Graybiel, A.M., and Agid, Y.A. (1988). Melanized dopaminergic neurons are differentially susceptible to degeneration in Parkinson's disease. Nature *334*, 345–348.

Hirvonen, J., and Palva, S. (2015). Cortical localization of phase and amplitude dynamics predicting access to somatosensory awareness. Hum. Brain Mapp. *326*, 311–326.

Hoffman, K.L., Dragan, M.C., Leonard, T.K., Micheli, C., Montefusco-Siegmund, R., and Valiante, T. a (2013). Saccades during visual exploration align hippocampal 3-8 Hz rhythms in human and non-human primates. Front. Syst. Neurosci. *7*, 43.

Honey, C.J., Thesen, T., Donner, T.H., Silbert, L.J., Carlson, C.E., Devinsky, O., Doyle, W.K., Rubin, N., Heeger, D.J., and Hasson, U. (2012). Slow cortical dynamics and the accumulation of information over long timescales. Neuron *76*, 423–434.

Hoogenboom, N., Schoffelen, J.M., Oostenveld, R., Parkes, L.M., and Fries, P. (2006). Localizing human visual gamma-band activity in frequency, time and space. NeuroImage *29*, 764–773.

Hoppensteadt, F.C., and Izhikevich, E.M. (1996). Synaptic organizations and dynamical properties of weakly connected neural oscillators. I. Analysis of a canonical model. Biol Cybern *75*, 117–127.

Hoppensteadt, F.C., and Izhikevich, E.M. (1998). Thalamo-cortical interactions modeled by weakly connected oscillators: could the brain use FM radio principles? Biosystems *48*, 85–94.

Hubel, D.H., and Wiesel, T.N. (1959). Receptive fields of single neurones in the cat's striate cortex. J. Physiol. *148*, 574–591.

Hubel, D.H., and Wiesel, T.N. (1960). Receptive fields of optic nerve fibres in the spider monkey. J. Physiol. *154*, 572–580.

Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. *160*, 106–154.2.

Huettel, S.A., Mack, P.B., and McCarthy, G. (2002). Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. Nat. Neurosci. *5*, 485–490.

Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., Ouden, H.E. den, and Stephan, K.E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron *80*, 519–530.

Ionta, S., Heydrich, L., Lenggenhager, B., Mouthon, M., Fornari, E., Chapuis, D., Gassert, R., and Blanke, O. (2011). Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective. Neuron *70*, 363–374.

Ionta, S., Martuzzi, R., Salomon, R., and Blanke, O. (2014). The brain network reflecting bodily self-consciousness: a functional connectivity study. Soc. Cogn. Affect. Neurosci.

Izard, V., Sann, C., Spelke, E.S., and Streri, A. (2009). Newborn infants perceive abstract numbers. Proc. Natl. Acad. Sci. *106*, 10382–10385.

Izhikevich, E.M. (2004). Which model to use for cortical spiking neurons? IEEE Trans. Neural Netw. *15*, 1063–1070.

Jadi, M.P., and Sejnowski, T.J. (2014). Cortical oscillations arise from contextual interactions that regulate sparse coding. Proc. Natl. Acad. Sci. U. S. A. *111*, 6780–6785.

James, G., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning: With Applications in R (Springer London, Limited).

Javoy-Agid, F., and Agid, Y. (1980). Is the mesocortical dopaminergic system involved in Parkinson disease? Neurology *30*, 1326–1330.

Jaynes, E.T. (2003). Probability Theory: The Logic of Science.

Jia, X., Xing, D., and Kohn, A. (2013). No consistent relationship between gamma power and peak frequency in macaque primary visual cortex.

Jolivet, R., Schuermann, F., Berger, T.K., Naud, R., Gerstner, W., and Roth, A. (2008). The quantitative single-neuron modeling competition. Biol. Cybern. *99*, 417–426.

Jones, S.R., Pritchett, D.L., Stufflebeam, S.M., Hämäläinen, M., and Moore, C.I. (2007). Neural correlates of tactile detection: a combined magnetoencephalography and biophysically based computational modeling study. J. Neurosci. Off. J. Soc. Neurosci. *27*, 10751–10764.

Jones, S.R., Pritchett, D.L., Sikora, M.A., Stufflebeam, S.M., Ha, M., and Moore, C.I. (2009). Quantitative Analysis and Biophysically Realistic Neural Modeling of the MEG Mu Rhythm : Rhythmogenesis and Modulation of Sensory-Evoked Responses. 3554–3572.

Joshi, S., Li, Y., Kalwani, R.M., and Gold, J.I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. Neuron *89*, 221–234.

Jutras, M.J., and Buffalo, E. a. (2010). Synchronous neural activity and memory formation. Curr. Opin. Neurobiol. *20*, 150–155.

Kahn, I., Yeshurun, Y., Rotshtein, P., Fried, I., Ben-Bashat, D., and Hendler, T. (2002). The role of the amygdala in signaling prospective outcome of choice. Neuron *33*, 983–994.

Kalman, R. (1959). On the general theory of control systems. IRE Trans. Autom. Control *4*, 110–110.

Kalman, R.E. (1963). Mathematical description of linear dynamical systems. J. Soc. Ind. Appl. Math. Ser. Control *1*, 152–192.

Kalman, R.E., and Bucy, R.S. (1961). New results in linear filtering and prediction theory. J. Basic Eng. *83*, 95–108.

Kalman, R.E., and others (1960). A new approach to linear filtering and prediction problems. J. Basic Eng. *82*, 35–45.

Kaplan, D. (2009). Structural Equation Modeling: Foundations and Extensions (SAGE Publications).

Kass, R.E., Eden, U., and Brown, E.N. (2014). Analysis of Neural Data (New York: Springer).

Kepecs, A., and Mainen, Z.F. (2012a). A computational framework for the study of confidence in humans and animals. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *367*, 1322–1337.

Kepecs, A., and Mainen, Z.F. (2012b). A computational framework for the study of confidence in humans and animals. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *367*, 1322–1337.

Kepecs, A., Uchida, N., Zariwala, H.A., and Mainen, Z.F. (2008). Neural correlates, computation and behavioural impact of decision confidence. Nature *455*, 227–231.

Van Kerkoerle, T., Self, M.W., Dagnino, B., Gariel-Mathis, M.A., Poort, J., van der Togt, C., and Roelfsema, P.R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc Natl Acad Sci U A *111*, 14332–14341.

Kiani, R., and Shadlen, M.N. (2009a). Representation of confidence associated with a decision by neurons in the parietal cortex. Science *324*, 759–764.

Kiani, R., and Shadlen, M.N. (2009b). Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex. Science *324*, 759–764.

Kiani, R., Corthell, L., and Shadlen, M.N. (2014). Choice Certainty Is Informed by Both Evidence and Decision Time. Neuron *84*, 1329–1342.

Kilner, J.M., and Lemon, R.N. (2013). What we know currently about mirror neurons. Curr Biol *23*, R1057–R1062.

Kiorpes, L., Dobkins, K., and Mendola, J.D. (2013). Linking hypotheses in visual neuroscience. Vis. Neurosci. *30*, 183–184.

Kish, S.J., Shannak, K., and Hornykiewicz, O. (1988). Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. Pathophysiologic and clinical implications. N. Engl. J. Med. *318*, 876–880.

Klein, I., Dubois, J., Mangin, J.F., Kherif, F., Flandin, G., Poline, J.B., Denis, M., Kosslyn, S.M., and Le Bihan, D. (2004). Retinotopic organization of visual mental images as revealed by functional magnetic resonance imaging. Brain Res Cogn Brain Res *22*, 26–31.

Klopp, J., Marinkovic, K., Chauvel, P., Nenov, V., and Halgren, E. (2000). Early widespread cortical distribution of coherent fusiform face selective activity. Hum. Brain Mapp. *11*, 286–293.

Knill, D.C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci. *27*, 712–719.

Kolossa, A., Fingscheidt, T., Wessel, K., and Kopp, B. (2013). A model-based approach to trial-by-trial p300 amplitude fluctuations. Front. Hum. Neurosci. *6*, 359.

König, P., and Schillen, T.B. (1991). Stimulus-Dependent Assembly Formation of Oscillatory Responses: I. Synchronization. Neural Comput. *3*, 155–166.

Krebs, R.M., Boehler, C.N., Roberts, K.C., Song, A.W., and Woldorff, M.G. (2012). The Involvement of the Dopaminergic Midbrain and Cortico-Striatal-Thalamic Circuits in the Integration of Reward Prospect and Attentional Task Demands. Cereb. Cortex *22*, 607–615.

Kuki, T., Fujihara, K., Miwa, H., Tamamaki, N., Yanagawa, Y., and Mushiake, H. (2015). Contribution of parvalbumin and somatostatin-expressing GABAergic neurons to slow oscillations and the balance in beta-gamma oscillations across cortical layers. Front. Neural Circuits *9*, 6.

Kuntimad, G., and Ranganath, H.S. (1999). Perfect image segmentation using pulse coupled neural networks. IEEE Trans. Neural Netw. Publ. IEEE Neural Netw. Counc. *10*, 591–598.

Kuzmina, M., Manykin, E., and Surina, I. (2004). Oscillatory network with self-organized dynamical connections for synchronization-based image segmentation. Biosystems *76*, 43–53.

Lachaux, J.P., Rodriguez, E., Martinerie, J., and Varela, F.J. (1999). Measuring phase synchrony in brain signals. Hum. Brain Mapp. *8*, 194–208.

Lachaux, J.P., Rodriguez, E., Martinerie, J., Adam, C., Hasboun, D., and Varela, F.J. (2000). A quantitative study of gamma-band activity in human intracranial recordings triggered by visual stimuli. Eur. J. Neurosci. *12*, 2608–2622.

Lachmann, T. (2002). Reading Disability as a Deficit in Functional Coordination. In Basic Functions of Language, Reading and Reading Disability, E. Witruk, A.D. Friederici, and T. Lachmann, eds. (Boston, MA: Springer US), pp. 165–198.

Lak, A., Costa, G.M., Romberg, E., Koulakov, A.A., Mainen, Z.F., and Kepecs, A. (2014). Orbitofrontal Cortex Is Required for Optimal Waiting Based on Decision Confidence. Neuron *84*, 190–201.

Lamme, V. a (1995). The neurophysiology of figure-ground segregation in primary visual cortex. J. Neurosci. Off. J. Soc. Neurosci. *15*, 1605–1615.

Lamme, V. a F., and Roelfsema, P.R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci. *23*, 571–579.

De Lange, F.P., Jensen, O., and Dehaene, S. (2010). Accumulation of evidence during sequential decision making: the importance of top-down factors. J. Neurosci. Off. J. Soc. Neurosci. *30*, 731–738.

De Lange, F.P., Rahnev, D.A., Donner, T.H., and Lau, H. (2013). Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. J. Neurosci. Off. J. Soc. Neurosci. *33*, 1400–1410.

LaRocque, J.J., Lewis-Peacock, J.A., Drysdale, A.T., Oberauer, K., and Postle, B.R. (2013). Decoding attended information in short-term memory: an EEG study. J. Cogn. Neurosci. *25*, 127–142.

Lee, S., and Jones, S.R. (2013). Distinguishing mechanisms of gamma frequency oscillations in human current source signals using a computational model of a laminar neocortical network. Front. Hum. Neurosci. *7*, 869.

Leenders, K.L., Palmer, A.J., Quinn, N., Clark, J.C., Firnau, G., Garnett, E.S., Nahmias, C., Jones, T., and Marsden, C.D. (1986). Brain dopamine metabolism in patients with Parkinson's disease measured with positron emission tomography. J. Neurol. Neurosurg. Psychiatry *49*, 853–860.

Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. Science *317*, 1096–1099.

Łęski, S., Lindén, H., Tetzlaff, T., Pettersen, K.H., and Einevoll, G.T. (2013). Frequency dependence of signal power and spatial reach of the local field potential. PLoS Comput. Biol. *9*, e1003137.

Lewis-Peacock, J.A., Drysdale, A.T., Oberauer, K., and Postle, B.R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. J. Cogn. Neurosci. *24*, 61–79.

Li, N., and DiCarlo, J.J. (2008). Unsupervised Natural Experience Rapidly Alters Invariant Object Representation in Visual Cortex. Science *321*, 1502–1507.

Lieder, F., Daunizeau, J., Garrido, M.I., Friston, K.J., and Stephan, K.E. (2013). Modelling trial-by-trial changes in the mismatch negativity. PLoS Comput. Biol. *9*, e1002911.

Lindén, H., Tetzlaff, T., Potjans, T.C., Pettersen, K.H., Grün, S., Diesmann, M., and Einevoll, G.T. (2011). Modeling the spatial reach of the LFP. Neuron *72*, 859–872.

Lisman, J.E., and Jensen, O. (2013). The ?-? neural code. Neuron *77*, 1002–1016.

Loewenfeld, I.E. (1993). The pupil: Anatomy, physiology, and clinical applications (Detroit: Wayne State University Press).

Lowet, E., Roberts, M.J., and de Weerd, P. (2012). Gamma frequency dependent horizontal interaction in macaque V1. In Society for Neuroscience Meeting, (New Orleans), p. 646.19/E33.

Lowet, E., Roberts, M., Hadjipapas, A., Peter, A., van der Eerden, J., and De Weerd, P. (2015a). Input-Dependent Frequency Modulation of Cortical Gamma Oscillations Shapes Spatial Synchronization and Enables Phase Coding. PLoS Comput Biol *11*, e1004072.

Lowet, E., Roberts, M.J., Bosman, C.A., Fries, P., and De Weerd, P. (2015b). Areas V1 and V2 show microsaccade-related 3–4-Hz covariation in gamma power and frequency. Eur. J. Neurosci. n/a – n/a.

Lowet, E., Roberts, M.J., Bonizzi, P., Karel, J., and De Weerd, P. (2016). Quantifying Neural Oscillatory Synchronization: A Comparison between Spectral Coherence and Phase-Locking Value Approaches. PLoS ONE *11*, e0146443.

Lund, J.S. (1988). Anatomical organization of macaque monkey striate visual cortex. Annu. Rev. Neurosci. *11*, 253–288.

Ma, W.J., and Jazayeri, M. (2014). Neural Coding of Uncertainty and Probability. Annu. Rev. Neurosci. *37*, 205–220.

Ma, W.J., Beck, J.M., Latham, P.E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. Nat. Neurosci. *9*, 1432–1438.

Macmillan, N.A., and Creelman, C.D. (1991). Detection Theory: A User's Guide. (New York: Cambridge University Press).

Maidhof, C. (2013). Error monitoring in musicians. Front. Hum. Neurosci. *7*, 401.

Maidhof, C., Pitkäniemi, A., and Tervaniemi, M. (2013). Predictive error detection in pianists: a combined ERP and motion capture study. Front. Hum. Neurosci. *7*, 587.

Maier, A., Adams, G.K., Aura, C., and Leopold, D. a (2010). Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. Front. Syst. Neurosci. *4*, 1–11.

Mainen, Z.F., and Kepecs, A. (2009). Neural representation of behavioral outcomes in the orbitofrontal cortex. Curr. Opin. Neurobiol. *19*, 84–91.

Makin, T.R., Scholz, J., Filippini, N., Henderson Slater, D., Tracey, I., and Johansen-Berg, H. (2013). Phantom pain is associated with preserved structure and function in the former hand area. Nat. Commun. *4*, 1570.

Von Der Malsburg, C. (1994). The correlation theory of brain function. Models Neural Netw. II 1–26.

Mandel, A., Bourguignon, M., Parkkonen, L., and Hari, R. (2016). Sensorimotor activation related to speaker vs. listener role during natural conversation. Neurosci. Lett. *614*, 99–104.

Marin, R.S. (1991). Apathy: a neuropsychiatric syndrome. J. Neuropsychiatry Clin. Neurosci. *3*, 243–254.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. Nat. Rev. Neurosci. *5*, 793–807.

Mars, R.B., Debener, S., Gladwin, T.E., Harrison, L.M., Haggard, P., Rothwell, J.C., and Bestmann, S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. J. Neurosci. Off. J. Soc. Neurosci. *28*, 12539–12545.

Marsolek, C.J. (2008). What antipriming reveals about priming. Trends Cogn. Sci. *12*, 176–181.

Masquelier, T.T., Hugues, E., Deco, G., and Thorpe, S.J. (2009). Oscillations, phase-of-firing coding, and spike timing-dependent plasticity: an efficient learning scheme. J. Neurosci. Off. J. Soc. Neurosci. *29*, 13484–13493.

May, A. (2011). Experience-dependent structural plasticity in the adult human brain. Trends Cogn. Sci. *15*, 475–482.

Mazzoni, A., Brunel, N., Cavallari, S., Logothetis, N.K., and Panzeri, S. (2011). Cortical dynamics during naturalistic sensory stimulations: experiments and models. J. Physiol. Paris *105*, 2–15.

Mazzoni, A., Lindén, H., Cuntz, H., Lansner, A., Panzeri, S., and Einevoll, G.T. (2015). Computing the Local Field Potential (LFP) from Integrate-and-Fire Network Models. PLoS Comput. Biol. *11*, e1004584.

Mazzoni, P., Hristova, A., and Krakauer, J.W. (2007). Why Don't We Move Faster? Parkinson's Disease, Movement Vigor, and Implicit Motivation. J. Neurosci. *27*, 7105–7116.

McGinley, M.J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C.R., Tolias, A.S., Cardin, J.A., and McCormick, D.A. (2015). Waking state: rapid variations modulate neural and behavioral responses. Neuron *87*, 1143–1161.

Meyniel, F., Sigman, M., and Mainen, Z.F. (2015a). Confidence as Bayesian Probability: From Neural Origins to Behavior. Neuron *88*, 78–92.

Meyniel, F., Schlunegger, D., and Dehaene, S. (2015b). The Sense of Confidence during Probabilistic Learning: A Normative Account. PLOS Comput. Biol. *11*, e1004305.

Miall, R.C., and Wolpert, D.M. (1996). Forward Models for Physiological Motor Control. Neural Netw. *9*, 1265–1279.

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.-M., Kennedy, H., and Fries, P. (2016). Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. Neuron *89*, 384–397.

Milner, P.M. (1974). A model for visual shape recognition. Psychol. Rev. *81*, 521–535.

Mink, J.W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. Prog. Neurobiol. *50*, 381–425.

Mises, R.V., and Pollaczek-Geiringer, H. (1929). Praktische Verfahren der Gleichungsauflösung . ZAMM - J. Appl. Math. Mech. Z. Für Angew. Math. Mech. *9*, 58–77.

Mogenson, G.J., Jones, D.L., and Yim, C.Y. (1980). From motivation to action: functional interface between the limbic system and the motor system. Prog. Neurobiol. *14*, 69–97.

Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. Science *319*, 1543–1546.

Moore, J.W., Lagnado, D., Deal, D.C., and Haggard, P. (2009). Feelings of control: Contingency determines experience of action. Cognition *110*, 279–283.

Morrison, S.E., and Salzman, C.D. (2010). Re-valuing the amygdala. Curr. Opin. Neurobiol. *20*, 221–230.

Moulton, E., Monzalvo, K., Bouhali, F., Hannagan, T., Thiebaut de Schotten, M., Lebenberg, J., Poupon, C., Hui Zhang, G., Dehaene, S., Dehaene-Labertz, G., et al. (submitted). White matter connections of fusiform areas: A longitudinal study in children learning to read. Hum. Brain Mapp.

Murakami, S., and Okada, Y. (2006). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. J. Physiol. *575*, 925–936.

Musall, S., von Pföstl, V., Rauch, A., Logothetis, N.K., and Whittingstall, K. (2014). Effects of neural synchrony on surface EEG. Cereb. Cortex N. Y. N 1991 *24*, 1045–1053.

Muthukumaraswamy, S.D., Singh, K.D., Swettenham, J.B., and Jones, D.K. (2010). Visual gamma oscillations and evoked responses: Variability, repeatability and structural MRI correlates. NeuroImage *49*, 3349–3357.

Ngo, H.-V.V., Martinetz, T., Born, J., and Mölle, M. (2013). Auditory Closed-Loop Stimulation of the Sleep Slow Oscillation Enhances Memory. Neuron *78*, 545–553.

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. Proc. Natl. Acad. Sci. U. S. A. *109*, 11860–11865.

Nieder, A. (2013). Coding of abstract quantity by "number neurons" of the primate brain. J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol. *199*, 1–16.

Nieder, A., and Merten, K. (2007). A Labeled-Line Code for Small and Large Numerosities in the Monkey Prefrontal Cortex. J. Neurosci. *27*, 5986–5993.

Nieuwenhuis, S., Aston-Jones, G., and Cohen, J.D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. Psychol. Bull. *131*, 510–532.

Nieuwenhuis, S., Geus, E.J.D., and Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. Psychophysiology *48*, 162–175.

Niv, Y., Daw, N.D., and Dayan, P. (2006). How fast to work: Response vigor, motivation and tonic dopamine. In Advances in Neural Information Processing Systems 18, Y. Weiss, B. Schölkopf, and J.C. Platt, eds. (MIT Press), pp. 1019–1026.

Nunez, P. (2006). Electric fields of the brain: the neurophysics of EEG.

Nunez, P.L., and Srinivasan, R. (2006a). A theoretical basis for standing and traveling brain waves measured with human EEG with implications for an integrated consciousness. Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. *117*, 2424–2435.

Nunez, P.L., and Srinivasan, R. (2006b). Electric Fields of the Brain (Oxford University Press).

O'Connell, R.G., Dockree, P.M., and Kelly, S.P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nat. Neurosci. *15*, 1729–1735.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 156869.

O'Reilly, J.X., Schüffelgen, U., Cuell, S.F., Behrens, T.E.J., Mars, R.B., and Rushworth, M.F.S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proc. Natl. Acad. Sci. U. S. A. *110*, E3660–E3669.

Ossmy, O., Moran, R., Pfeffer, T., Tsetsos, K., Usher, M., and Donner, T.H. (2013). The timescale of perceptual evidence integration can be adapted to the environment. Curr. Biol. *23*, 981–986.

Packard, M.G., and McGaugh, J.L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiol. Learn. Mem. *65*, 65–72.

Pallier, C., Devauchelle, A.-D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. Proc. Natl. Acad. Sci. U. S. A. *108*, 2522–2527.

Palva, S., and Palva, J.M. (2012). Discovering oscillatory interaction networks with M/EEG: Challenges and breakthroughs. Trends Cogn. Sci. *16*, 219–229.

Pan, Y., Lin, B., Zhao, Y., and Soto, D. (2014). Working memory biasing of visual perception without awareness. Atten. Percept. Psychophys. *76*, 2051–2062.

Pasquereau, B., and Turner, R.S. (2013). Limited encoding of effort by dopamine neurons in a cost-benefit trade-off task. J. Neurosci. Off. J. Soc. Neurosci. *33*, 8288–8300.

Patla, A.E. (1997). Understanding the roles of vision in the control of human locomotion. Gait Posture *5*, 54–69.

Patla, A.E., Rietdyk, S., Martin, C., and Prentice, S. (1996). Locomotor Patterns of the Leading and the Trailing Limbs as Solid and Fragile Obstacles Are Stepped Over: Some Insights Into the Role of Vision During Locomotion. J. Mot. Behav. *28*, 35–47.

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., and O'Doherty, J.P. (2013). The neural representation of unexpected uncertainty during value-based decision making. Neuron *79*, 191–201.

Pearce, J.M., Roberts, A.D., and Good, M. (1998). Hippocampal lesions disrupt navigation based on cognitive maps but not heading vectors. Nature *396*, 75–77.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.).

Van Pelt, S., Boomsma, D.I., and Fries, P. (2012). Magnetoencephalography in Twins Reveals a Strong Genetic Determination of the Peak Frequency of Visually Induced Gamma-Band Synchronization. J. Neurosci. *32*, 3388–3392.

Perea, G., Sur, M., and Araque, A. (2014). Neuron-glia networks: integral gear of brain function. Front. Cell. Neurosci. *8*, 378.

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

Pessiglione, M., Guehl, D., Rolland, A.-S., François, C., Hirsch, E.C., Féger, J., and Tremblay, L. (2005). Thalamic neuronal activity in dopamine-depleted primates: evidence for a loss of functional segregation within basal ganglia circuits. J. Neurosci. Off. J. Soc. Neurosci. *25*, 1523–1531.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., and Frith, C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature *442*, 1042–1045.

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R.J., and Frith, C.D. (2007). How the brain translates money into force: a neuroimaging study of subliminal motivation. Science *316*, 904–906.

Peterson, M.A., and Salvagio, E. (2009). Attention and competition in figure-ground perception. Prog. Brain Res. *176*, 1–13.

Pettersen, K.H., Lindén, H., Tetzlaff, T., and Einevoll, G.T. (2014). Power laws from linear neuronal cable theory: power spectral densities of the soma potential, soma membrane current and single-neuron contribution to the EEG. PLoS Comput. Biol. *10*, e1003928.

Pikovsky, A., Rosenblum, M., Kurths, J., and Hilborn, R.C. (2002). Synchronization: A Universal Concept in Nonlinear Science. Am. J. Phys. *70*, 655.

Polack, P.-O., Friedman, J., and Golshani, P. (2013). Cellular mechanisms of brain state-dependent gain modulation in visual cortex. Nat. Neurosci. *16*, 1331–1339.

Polyn, S.M., Natu, V.S., Cohen, J.D., and Norman, K.A. (2005). Category-specific cortical activity precedes retrieval during memory search. Science *310*, 1963–1966.

Poort, J., Raudies, F., Wannig, A., Lamme, V.A.F., Neumann, H., and Roelfsema, P.R. (2012). The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. Neuron *75*, 143–156.

Posner, M.I., and Gilbert, C.D. (1999). Attention and primary visual cortex. Proc. Natl. Acad. Sci. U. S. A. *96*, 2585–2587.

Potjans, T.C., and Diesmann, M. (2014). The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. Cereb. Cortex N. Y. N 1991 *24*, 785–806.

Prinz, W. (1997). Perception and action planning. Eur. J. Cogn. Psychol. *9*, 129–154.

Quarteroni, A., Sacco, R., and Saleri, F. (2007). Numerical Mathematics (Berlin, Heidelberg: Springer Berlin Heidelberg).

Ramachandran, V.S. (1988). PERCEPTION OF SHAPE FROM SHADING. Nature *331*, 163–166.

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

Ray, S., and Maunsell, J.H.R. (2010). Differences in Gamma Frequencies across Visual Cortex Restrict Their Possible Use in Computation. Neuron *67*, 885–896.

Ray, S., and Maunsell, J.H.R. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol. *9*, e1000610.

Ray, S., Ni, A.M., and Maunsell, J.H.R. (2013). Strength of gamma rhythm depends on normalization. PLoS Biol. *11*, e1001477.

Reed, K.E., and Bidner, L.R. (2004). Primate communities: past, present, and possible future. Am. J. Phys. Anthropol. *Suppl 39*, 2–39.

Remy, P., Doder, M., Lees, A., Turjanski, N., and Brooks, D. (2005). Depression in Parkinson's disease: loss of dopamine and noradrenaline innervation in the limbic system. Brain J. Neurol. *128*, 1314–1322.

Rescorla, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Classical Conditioning II: Current Research and Theory, (New York: Appleton Century Crofts), pp. 64–99.

Reynolds, J.H., and Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. Neuron *24*, 19–29.

Rieke, F., Warland, D., De Ruyter Van Steveninck, R., and Bialek, W. (1997). Spikes: Exploring the Neural Code.

Rigoux, L., and Guigon, E. (2012). A model of reward- and effort-based optimal decision making and motor control. PLoS Comput. Biol. *8*, e1002716.

Rizzolatti, G., and Luppino, G. (2001). The cortical motor system. Neuron *31*, 889–901.

Roberts, M.J., Lowet, E., Brunet, N.M., TerWal, M., Tiesinga, P., Fries, P., and DeWeerd, P. (2013). Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. Neuron *78*, 523–536.

Rodriguez-Oroz, M.C., Jahanshahi, M., Krack, P., Litvan, I., Macias, R., Bezard, E., and Obeso, J.A. (2009). Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. Lancet Neurol. *8*, 1128–1139.

Roelfsema, P.R., Engel, A.K., Konig, P., and Singer, W. (1997). Visuomotor integration is associated with zero time-lag sychronization among cortical areas. Nature *385*, 157–161.

Roelfsema, P.R., Ooyen, A. van, and Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. Trends Cogn. Sci. *14*, 64–71.

Roesch, M.R., Calu, D.J., and Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat. Neurosci. *10*, 1615–1624.

Rubehn, B., Bosman, C., Oostenveld, R., Fries, P., and Stieglitz, T. (2009). A MEMS-based flexible multichannel ECoG-electrode array. J Neural Eng *6*, 36003–36010.

Sakurai, Y. (1996). Population coding by cell assemblies--what it really is in the brain. Neurosci Res *26*, 1–16.

Salamone, J.D., and Correa, M. (2012). The Mysterious Motivational Functions of Mesolimbic Dopamine. Neuron *76*, 470–485.

Salamone, J.D., Correa, M., Nunes, E.J., Randall, P.A., and Pardo, M. (2012). The behavioral pharmacology of effort-related choice behavior: dopamine, adenosine and beyond. J. Exp. Anal. Behav. *97*, 125–146.

Salinas, E., and Sejnowski, T.J. (2001). Correlated neuronal activity and the flow of neural information. Nat. Rev. Neurosci. *2*, 539–550.

Sandberg, K., Timmermans, B., Overgaard, M., and Cleeremans, A. (2010). Measuring consciousness: is one measure better than the other? Conscious. Cogn. *19*, 1069–1078.

Santos, F.J., Oliveira, R.F., Jin, X., and Costa, R.M. (2015). Corticostriatal dynamics encode the refinement of specific behavioral variability during skill learning. eLife *4*, e09423.

Sara, S.J. (2009). The locus coeruleus and noradrenergic modulation of cognition. Nat. Rev. Neurosci. *10*, 211–223.

Sarter, M., Parikh, V., and Howe, W.M. (2009). Phasic acetylcholine release and the volume transmission hypothesis: time to move on. Nat. Rev. Neurosci. *10*, 383–390.

Schellenberger Costa, M., Weigenand, A., Ngo, H.-V.V., Marshall, L., Born, J., Martinetz, T., and Claussen, J.C. (In review). A thalamocortical neural mass model of the EEG during NREM sleep and its response to auditory stimulation. PLOS Comput. Biol.

Schellenberger Costa, M., Born, J., Claussen, J.C., and Martinetz, T. (In print). Modeling the effect of sleep regulation on a neural mass model. J. Comput. Neurosci.

Schmidt, L., d' Arc, B.F., Lafargue, G., Galanaud, D., Czernecki, V., Grabli, D., Schüpbach, M., Hartmann, A., Lévy, R., Dubois, B., et al. (2008). Disconnecting force from money: effects of basal ganglia damage on incentive motivation. Brain J. Neurol. *131*, 1303–1310.

Schmidt, L., Lebreton, M., Cléry-Melin, M.-L., Daunizeau, J., and Pessiglione, M. (2012). Neural mechanisms underlying motivation of mental versus physical effort. PLoS Biol. *10*, e1001266.

Schouppe, N., Demanet, J., Boehler, C.N., Ridderinkhof, K.R., and Notebaert, W. (2014). The Role of the Striatum in Effort-Based Decision-Making in the Absence of Reward. J. Neurosci. *34*, 2148–2154.

Schroeder, C.E., Wilson, D. a., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of Active Sensing and perceptual selection. Curr. Opin. Neurobiol. *20*, 172–176.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A Neural Substrate of Prediction and Reward. Science *275*, 1593–1599.

Schwemmer, M.A., and Lewis, T.J. (2012). Phase Response Curves in Neuroscience. In Phase Response Curves in Neuroscience, pp. 3–31.

Self, M.W., van Kerkoerle, T., Supèr, H., and Roelfsema, P.R. (2013). Distinct Roles of the Cortical Layers of Area V1 in Figure-Ground Segregation. Curr. Biol.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., and LeCun, Y. (2014). OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. Int. Conf. Learn. Represent. ICLR 2014.

Shadmehr, R., Xivry, J.J.O. de, Xu-Wilson, M., and Shih, T.-Y. (2010). Temporal Discounting of Reward and the Cost of Time in Motor Control. J. Neurosci. *30*, 10507–10516.

Shafritz, K.M., Gore, J.C., and Marois, R. (2002). The role of the parietal cortex in visual feature binding. Proc. Natl. Acad. Sci. U. S. A. *99*, 10917–10922.

Shannon, C.E. (1948). A Mathematical Theory of Communication. Bell Syst. Tech. J. *27*, 379–423.

Shepard, R.N., and Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. Science *171*, 701–703.

Shepard, R.N., Kilpatric, D.W., and Cunningham, J.P. (1975). The internal representation of numbers. Cognit. Psychol. *7*, 82–138.

Shibata, E., Sasaki, M., Tohyama, K., Kanbara, Y., Otsuka, K., Ehara, S., and Sakai, A. (2006). Age-related changes in locus ceruleus on neuromelanin magnetic resonance imaging at 3 Tesla. Magn. Reson. Med. Sci. *5*, 197–200.

Shima, K., Isoda, M., Mushiake, H., and Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. Nature *445*, 315–318.

Shum, J., Hermes, D., Foster, B.L., Dastjerdi, M., Rangarajan, V., Winawer, J., Miller, K.J., and Parvizi, J. (2013). A Brain Area for Visual Numerals. J. Neurosci. Off. J. Soc. Neurosci. *33*, 6709–6715.

Siegel, M., Engel, A.K., and Donner, T.H. (2011). Cortical network dynamics of perceptual decision-making in the human brain. Front. Hum. Neurosci. *5*.

Siegel, M., Donner, T.H., and Engel, A.K. (2012). Spectral fingerprints of large-scale neuronal interactions. Nat. Rev. Neurosci. *13*, 121–134.

Silani, G., Lamm, C., Ruff, C.C., and Singer, T. (2013). Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. J. Neurosci. Off. J. Soc. Neurosci. *33*, 15466–15476.

Sincich, L.C., and Horton, J.C. (2005). THE CIRCUITRY OF V1 AND V2: Integration of Color, Form, and Motion. Annu. Rev. Neurosci. *28*, 303–326.

Singer, W. (1995). Visual Feature Integration and the Temporal Correlation Hypothesis. Annu. Rev. Neurosci. *18*, 555–586.

Singer, W. (1999). Time as coding space? Curr. Opin. Neurobiol. *9*, 189–194.

Skvortsova, V., Palminteri, S., and Pessiglione, M. (2014). Learning To Minimize Efforts versus Maximizing Rewards: Computational Principles and Neural Correlates. J. Neurosci. *34*, 15621–15630.

Soto, D., and Silvanto, J. (2014). Reappraising the relationship between working memory and conscious awareness. Trends Cogn. Sci. *18*, 520–525.

De Sousa, A.A., Sherwood, C.C., Schleicher, A., Amunts, K., MacLeod, C.E., Hof, P.R., and Zilles, K. (2010). Comparative cytoarchitectural analyses of striate and extrastriate areas in hominoids. Cereb. Cortex *20*, 966–981.

Squires, K.C., Wickens, C., Squires, N.K., and Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. Science *193*, 1142–1146.

Starkstein, S.E., Mayberg, H.S., Preziosi, T.J., Andrezejewski, P., Leiguarda, R., and Robinson, R.G. (1992). Reliability, validity, and clinical correlates of apathy in Parkinson's disease. J. Neuropsychiatry Clin. Neurosci. *4*, 134–139.

Von Stein, a, Chiang, C., and König, P. (2000). Top-down processing mediated by interareal synchronization. Proc. Natl. Acad. Sci. U. S. A. *97*, 14748–14753.

Stephens, D.W., and Krebs, J.R. (1986). Foraging Theory (Princeton University Press).

Steriade, M. (2000). Corticothalamic resonance, states of vigilance and mentation. Neuroscience *101*, 243–276.

Stettler, D.D., Das, A., Bennett, J., and Gilbert, C.D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. Neuron *36*, 739–750.

Stevens, S.S. (1957). On the psychophysical law. Psychol. Rev. *64*, 153–181.

Stoianov, I., and Zorzi, M. (2012). Emergence of a "visual number sense" in hierarchical generative models. Nat. Neurosci. *15*, 194–196.

St Onge, J.R., and Floresco, S.B. (2009). Dopaminergic modulation of risk-based decision making. Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol. *34*, 681–697.

Strange, B.A., Duggins, A., Penny, W., Dolan, R.J., and Friston, K.J. (2005). Information theory, novelty and hippocampal responses: unpredicted or unpredictable? Neural Netw. Off. J. Int. Neural Netw. Soc. *18*, 225–230.

Strauss, M., Sitt, J.D., King, J.-R., Elbaz, M., Azizi, L., Buiatti, M., Naccache, L., van Wassenhove, V., and Dehaene, S. (2015). Disruption of hierarchical predictive coding during sleep. Proc. Natl. Acad. Sci. U. S. A. *112*, E1353–E1362.

Supèr, H., Spekreijse, H., and Lamme, V.A. (2001). A neural correlate of working memory in the monkey primary visual cortex. Science *293*, 120–124.

Sutton, R. (1992). Gain Adaptation Beats Least Squares? Proc. 7th Yale Workshop Adapt. Learn. Syst. 161–166.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement Learning (Cambridge, MA: MIT Press).

Swettenham, J.B., Muthukumaraswamy, S.D., and Singh, K.D. (2009). Spectral properties of induced and evoked gamma oscillations in human early visual cortex to moving and stationary stimuli. J. Neurophysiol. *102*, 1241–1253.

Tallon-Baudry, C., Bertrand, O., and Fischer, C. (2001). Oscillatory synchrony between human extrastriate areas during visual short-term memory maintenance. J. Neurosci. *21*, RC177–RC177.

Tarantola, A. (2005). Inverse Problem Theory and Methods for Model Parameter Estimation (Society for Industrial and Applied Mathematics).

Teller, D.Y. (1984). Linking propositions. Vision Res. *24*, 1233–1246.

Thobois, S., Ardouin, C., Lhommée, E., Klinger, H., Lagrange, C., Xie, J., Fraix, V., Coelho Braga, M.C., Hassani, R., Kistner, A., et al. (2010). Non-motor dopamine withdrawal syndrome after surgery for Parkinson's disease: predictors and underlying mesolimbic denervation. Brain J. Neurol. *133*, 1111–1127.

Thomson, A.M., West, D.C., Wang, Y., and Bannister, A.P. (2002). Synaptic Connections and Small Circuits Involving Excitatory and Inhibitory Neurons in Layers 2 – 5 of Adult Rat and Cat Neocortex : Triple Intracellular Recordings and Biocytin Labelling In Vitro. 936–953.

Tiesinga, P.H., and Sejnowski, T.J. (2010). Mechanisms for Phase Shifting in Cortical Networks and their Role in Communication through Coherence. *4*, 196.

Tiesinga, P., and Sejnowski, T.J. (2009). Cortical Enlightenment: Are Attentional Gamma Oscillations Driven by ING or PING? Neuron *63*, 727–732.

Tiesinga, P.H., Fellous, J.-M., Salinas, E., José, J. V, and Sejnowski, T.J. (2005). Inhibitory synchrony as a mechanism for attentional gain modulation. J. Physiol. Paris *98*, 296–314.

Tiesinga, P.H.E., Fellous, J.M., José, J. V, and Sejnowski, T.J. (2002). Information transfer in entrained cortical neurons. Network *13*, 41–66.

Tobler, P.N., Fiorillo, C.D., and Schultz, W. (2005). Adaptive Coding of Reward Value by Dopamine Neurons. Science *307*, 1642–1645.

Todorov, E., and Jordan, M.I. (2002). Optimal feedback control as a theory of motor coordination. Nat. Neurosci. *5*, 1226–1235.

Tomassini, A., Spinelli, D., Jacono, M., Sandini, G., and Morrone, M.C. (2015). Rhythmic Oscillations of Visual Contrast Sensitivity Synchronized with Action. J. Neurosci. *35*, 7019–7029.

Traub, R.D., Whittington, M. a, Colling, S.B., Buzsáki, G., and Jefferys, J.G. (1996). Analysis of gamma rhythms in the rat hippocampus in vitro and in vivo. J. Physiol. *493 (Pt 2*, 471–484.

Treadway, M.T., Buckholtz, J.W., Cowan, R.L., Woodward, N.D., Li, R., Ansari, M.S., Baldwin, R.M., Schwartzman, A.N., Kessler, R.M., and Zald, D.H. (2012). Dopaminergic Mechanisms of Individual Differences in Human Effort-Based Decision-Making. J. Neurosci. *32*, 6170–6176.

Treisman, a (1998). Feature binding, attention and object perception. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *353*, 1295–1306.

Treisman, A. (2004). Feature binding, attention, and object perception. In Essential Sources in the Scientific Study of Consciousness, p. 226.

Treisman, A.M., and Gelade, G. (1980). A feature-integration theory of attention. Cognit. Psychol. *12*, 97–136.

Tsai, H.-C., Zhang, F., Adamantidis, A., Stuber, G.D., Bonci, A., de Lecea, L., and Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science *324*, 1080–1084.

Tsetsos, K., Chater, N., and Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. Proc. Natl. Acad. Sci. U. S. A. *109*, 9659–9664.

Tsuchiya, N., and Koch, C. (2005). Continuous flash suppression reduces negative afterimages. Nat. Neurosci. *8*, 1096–1101.

Tsukada, M., Ichinose, N., Aihara, K., Ito, H., and Fujii, H. (1996). Dynamical Cell Assembly Hypothesis - Theoretical Possibility of Spatio-temporal Coding in the Cortex. Neural Netw *9*, 1303–1350.

Uhrig, L., Dehaene, S., and Jarraya, B. (2014). A hierarchy of responses to auditory regularities in the macaque brain. J. Neurosci. *34*, 1127–1132.

Usher, M., and McClelland, J.L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. Psychol. Rev. *108*, 550.

Vanrie, J., Dekeyser, M., and Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-light walkers. Perception *33*, 547–560.

VanRullen, R., Guyonneau, R., and Thorpe, S.J. (2005). Spike times make sense. Trends Neurosci *28*, 1–4.

Varazzani, C., San-Galli, A., Gilardeau, S., and Bouret, S. (2015). Noradrenaline and Dopamine Neurons in the Reward/Effort Trade-Off: A Direct Electrophysiological Comparison in Behaving Monkeys. J. Neurosci. *35*, 7866–7877.

Varela, F., Lachaux, J., Rodriguez, E., and Martinerie, J. (2001). the Brainweb : Phase Large-Scale Integration. *2*.

Verguts, T., and Fias, W. (2004). Representation of number in animals and humans: a neural model. J. Cogn. Neurosci. *16*, 1493–1504.

Verguts, T., Vassena, E., and Silvetti, M. (2015). Adaptive effort investment in cognitive and physical tasks: a neurocomputational model. Front. Behav. Neurosci. *9*, 57.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., and Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc Natl Acad Sci U A *108*, 20754–20759.

Wacongne, C., Changeux, J.-P., and Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. J. Neurosci. *32*, 3665–3678.

Walton, M.E., Kennerley, S.W., Bannerman, D.M., Phillips, P.E.M., and Rushworth, M.F.S. (2006). Weighing up the benefits of work: behavioral and neural analyses of effort-related decision making. Neural Netw. Off. J. Int. Neural Netw. Soc. *19*, 1302–1314.

Walton, M.E., Groves, J., Jennings, K.A., Croxson, P.L., Sharp, T., Rushworth, M.F.S., and Bannerman, D.M. (2009). Comparing the role of the anterior cingulate cortex and 6-hydroxydopamine nucleus accumbens lesions on operant effort-based decision making. Eur. J. Neurosci. *29*, 1678–1691.

Wang, X. (2010). Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. 1195–1268.

Wang, X.-J. (2008). Decision making in recurrent neuronal circuits. Neuron *60*, 215–234.

Wang, D., and Terman, D. (1995). Locally excitatory globally inhibitory oscillator networks. IEEE Trans. Neural Netw. Publ. IEEE Neural Netw. Counc. *6*, 283–286.

Wang, D., and Terman, D. (1997). Image Segmentation Based on Oscillatory Correlation. Neural Comput. *9*, 805–836.

Wardle, M.C., Treadway, M.T., Mayo, L.M., Zald, D.H., and de Wit, H. (2011). Amping up effort: effects of d-amphetamine on human effort-based decision-making. J. Neurosci. Off. J. Soc. Neurosci. *31*, 16597–16602.

Warren, R.M., and Warren, R.P. (1968). Helmholtz on Perception: Its Physiology and Development (John Wiley & Sons Inc).

Watkins, C.J.C.H. (1989). Learning from Delayed Rewards. King's College.

Weigenand, A., Schellenberger Costa, M., Ngo, H.-V.V., Claussen, J.C., and Martinetz, T. (2014). Characterization of k-complexes and slow wave activity in a neural mass model. PLoS Comput. Biol. *10*, e1003923.

Whittington, M. a, Traub, R.D., and Jefferys, J.G. (1995). Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. Nature *373*, 612–615.

Whittington, M. a, Traub, R.D., Kopell, N., Ermentrout, B., and Buhl, E.H. (2000). Inhibition-based rhythms: experimental and mathematical observations on network dynamics. Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol. *38*, 315–336.

Wiecki, T.V., Sofer, I., and Frank, M.J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. Front. Neuroinformatics *7*, 14.

Wildie, M., and Shanahan, M. (2011). Establishing Communication between Neuronal Populations through Competitive Entrainment. Front. Comput. Neurosci. *5*, 62.

Wilf, M., Strappini, F., Golan, T., Hahamy, A., Harel, M., and Malach, R. (2015). Spontaneously Emerging Patterns in Human Visual Cortex Reflect Responses to Naturalistic Sensory Stimuli. Cereb. Cortex N. Y. N 1991.

Willner, P., Phillips, G., Muscat, R., and Hood, P. (1992). Behavioural tests of the dopamine depletion hypothesis of neuroleptic-induced response decrement. Psychopharmacology (Berl.) *106*, 543–549.

Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., Singer, W., Desimone, R., Engel, A.K., and Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. Science *316*, 1609–1612.

Womelsdorf, T., Valiante, T. a, Sahin, N.T., Miller, K.J., and Tiesinga, P. (2014). Dynamic circuit motifs underlying rhythmic gain control, gating and integration. Nat. Neurosci. *17*, 1031–1039.

Wyart, V., Gardelle, V.D., Scholl, J., and Summerfield, C. (2012). Rhythmic fluctuations in evidence accumulation during decision making in the human brain. Neuron *76*, 847–858.

Xie, X., and Giese, M.A. (2002). Nonlinear dynamics of direction-selective recurrent neural media. Phys Rev E Stat Nonlin Soft Matter Phys *65*, 051904.

Xing, D., Shen, Y., Burns, S., Yeh, C.-I., Shapley, R., and Li, W. (2012). Stochastic Generation of Gamma-Band Activity in Primary Visual Cortex of Awake and Anesthetized Monkeys. J. Neurosci. *32*, 13873–13880.

Ben-Yakov, A., and Dudai, Y. (2011). Constructing Realistic Engrams: Poststimulus Activity of Hippocampus and Dorsal Striatum Predicts Subsequent Episodic Memory. J. Neurosci. *31*, 9032–9042.

Ben-Yakov, A., Eshel, N., and Dudai, Y. (2013). Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. J. Exp. Psychol. Gen. *142*, 1255–1263.

Ben-Yakov, A., Rubinson, M., and Dudai, Y. (2014). Shifting Gears in Hippocampus: Temporal Dissociation between Familiarity and Novelty Signatures in a Single Event. J. Neurosci. *34*, 12973–12981.

Yang, T., and Shadlen, M.N. (2007). Probabilistic reasoning by neurons. Nature *447*, 1075–1080.

Yoshioka, T., Blasdel, G.G., Levitt, J.B., and Lund, J.S. (1996). Relation between patterns of intrinsic lateral connectivity, ocular dominance, and cytochrome oxidase-reactive regions in macaque monkey striate cortex. Cereb. Cortex *6*, 297–310.

Zachariou, M., Roberts, M., Lowet, E., de Weerd, P., and Hadjipapas, A. (2015). Contrast-dependent modulation of gamma rhythm in v1: a network model. BMC Neurosci. *16*, O10.

Zandvakili, A., and Kohn, A. (2016). Coordinated Neuronal Activity Enhances Corticocortical Communication. Neuron *87*, 827–839.

Zhdanov, A., Nurminen, J., Baess, P., Hirvenkari, L., Jousmäki, V., Mäkelä, J.P., Mandel, A., Meronen, L., Hari, R., and Parkkonen, L. (2015). An Internet-Based Real-Time Audiovisual Link for Dual MEG Recordings. PloS One *10*, e0128485.

Zhigalov, A., Arnulfo, G., Nobili, L., Palva, S., and Palva, J.M. (2015). Relationship of Fast- and Slow-Timescale Neuronal Dynamics in Human MEG and SEEG. J. Neurosci. *35*, 5385–5396.

Zhou, G., Bourguignon, M., Parkkonen, L., and Hari, R. (2016). Neural signatures of hand kinematics in leaders vs. followers: A dual-MEG study. NeuroImage *125*, 731–738.

Zylberberg, A., Barttfeld, P., and Sigman, M. (2012). The construction of confidence in a perceptual decision. Front. Integr. Neurosci. *6*, 79.

Annex B: Dataset Information Cards

Task	Partner	Data / model name	DIC name	DIC registered	Link
3.1.1	ESI	Large-scale recordings from distributed and local visual networks during rest	Spontaneous activity in anesthetized cat area 17	yes	
3.1.1	CEA	MEEG recordings of the time course and manipulation of view-specific and view-independent models of objects	M/EEG Visual Stimulation in Humans (RSVP, rotating objects and mental rotation) - raw data	yes	http://sp3.s3.data.kit.edu/3_1_1/MEG_PredictiveVisualInternalModel/Raw/
3.1.1	WIS	The patterns of co-activation during natural sensory processing uncovered through resting state and naturalistic stimulation paradigms	Architecture of functional visual cognitive networks of the human brain	yes	
3.1.2	EKUT	Neurodynamical model for multistability and the perceptual organization effects including multiple views in action recognition Giese et al 2014	Perception Action - Codes	yes	http://sp3.s3.data.kit.edu/3.1.2/
3.1.2	EKUT	Model for new pathway that accounts for the influence of shading cues on action perception Publi. In prep.	Not available yet (extra data)		
3.1.2	EKUT	Spiking neuron model for a key circuit linking visual and motor representation of actions	Not available yet (extra data)		
3.1.3	EPFL	Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective	Neural correlates of self-location and first-person perspective	yes	
3.1.3	UB	The Study of Body Ownership and Agency Using Immersive Virtual Reality Methods	Understanding how body perception becomes a reference point for the sense of self - body ownership and agency	yes	
3.1.4	UM	Pyramidal Interneuron Network Gamma (PING) models of excitatory (E) and inhibitory (I) cells, monkey, model and	Models of gamma oscillations in visual cortex	yes	

		simulation results and A proof-of-principle layer-extended column model			
3.1.5	UH, AMU	Map of human inter-areal connectivity and phase lags based on resting-state SEEG	Human connectome of phase lags	yes	http://sp3.s3.data.kit.edu/3.1.5/
3.1.5	UH, AMU	Effects of multimodal distribution of delays in brain network dynamics, reduced model	pending		
3.2.1	CEA	Confidence during probabilistic reasoning, behavioural and fMRI recordings	Human networks involved in confidence (fMRI and behavior)	yes	Study1: http://s3.data.kit.edu/SP3/3_2_1/Study1_PerceptualConfidence Study2: data are available as on-line supplementary material of the Plos Computional Biology publication; they were also deposited on a server at http://s3.data.kit.edu/SP3/3_2_1/Study2_ProbabilistLearning_behavior Study3: data were deposited on a server at http://s3.data.kit.edu/SP3/3_2_1/Study3_ProbabilisticLearning_fMRI
3.2.1	FCHAMP	Confidence estimation on motor skill performance in mice	Confidence estimation on motor skill performance in mice	yes	
3.2.2	UPMC	Pharmacological manipulation of motivational processes	pending		
3.2.3	UvA	Pupil-linked brainstem responses and the computation of yes vs. no decisions (fMRI), brain stem modulation	Brainstem modulation of decision processes (human behaviour and fMRI)	yes	http://s3.data.kit.edu/SP3/3_2_3/Study1_yesno_fMRI
3.2.3	UKE	Pupil-linked modulation of the cortical dynamics underlying yes vs. no decisions (MEG)	Brainstem modulation of decision processes (human behaviour and MEG)	yes	http://s3.data.kit.edu/SP3/3_2_3/Study2_yesno_MEG

3.2.4	TASMC	Intracranial single cell and LFP dataset, decision making, human	intracranial recordings in motivational paradigm	yes	http://fmri-tlv.org/tomer.html
3.3.1	UHAIFA	Short-term cortical modulation by task repetition as signatures of procedural memory consolidation	pending		https://openfmri.org/dataset/ds000170/
3.3.2	WIS	Cognitive architecture of the initation of systems consolidation	Consolidation of realistic episodic memories - stage 1 and 2	yes	http://sp3.s3.data.kit.edu/3_3_2/fMRI/Consolidation_of_realistic_episodic_memories_stage_1/ http://sp3.s3.data.kit.edu/3_3_2/fMRI/Consolidation_of_realistic_episodic_memories_stage_2/
3.3.2	WIS	Brain activity that predicts episodic memory for brief narrative movie clips	Prestimulus predictors of memory encoding,	yes	http://sp3.s3.data.kit.edu/3_3_2/fMRI/Prestimulus_predictors_of_memory_encoding/
3.3.2	EKUT	Neural mass models of the sleeping brain	pending		
3.3.3	UMU	Short-term maintenance of conscious and non-conscious information	Non-conscious short-term memory	yes	http://sp3.s3.data.kit.edu/3_3_3/fMRI_raw_data/
3.4.1	UCL	Model of spatial navigation and spatial memory	Rat navigation simulation	yes	http://se/data/kit/edu/SP3/3.4.1/Simulation Results
3.5.1	CNRS	Emergence and self-organization of internal knowledge, tackling issues related to local vs. global feature processing,	Recordings from primary visual cortex of anaesthetised cat during visual stimulation	yes	https://hbp.unic.cnrs-gif.fr/db
3.5.2	CNRS	Data set precisely consists in two-photon calcium imaging of mouse V1 and A1 activity during time-varying auditory visual stimulation	Multimodal activity in visual cortex V1 of the mouse	yes	http://sp3.s3.data.kit.edu/3.5.2/Auditory_visual_dataset/
3.6.1	CEA	Cognitive architecture for the emergence of symbol-related areas: A dataset containing the cortical simulations from OverFeat for different geometric transforms of faces, tools, houses, letters, strings of letters and strings of pseudoletters	Cortical simulations for symbolic and non-symbolic stimuli	yes	http://s3.data.kit.edu/3.6.1

3.6.2	CEA	Encoding of syntactic structures: model of the emergence of human areas responsive to letter an number symbols	Encoding of syntactic structures	yes	http://sp3.s3.data.kit.edu/3.6.2
3.6.2	CEA	Bayesian Modeling of Expectation Effects in Sequences	Cortical encoding of probabilistic sequences (MEG and behavior)	yes	http://s3.data.kit.edu/SP3/3_6_2/Study_ExpectationBayesianModeling_MEG
3.6.2	CEA	Encoding of temporal structure by human and non-human primates	pending		
3.6.3	AALTO	The social brain, two person interactions	fMRI localizer of social brain	yes	http://ani.aalto.fi/en/ami_centre/

Annex C: Published reviews