Project Number:	284941	Project Title:	Human Brain Project

Document Title:	Implementation and validation of functional mapping methods
Document Filename:	SP03_D3.7.2
Deliverable Number:	D3.7.2
Deliverable Type:	Report
Work Package(s):	WP3.1, WP3.2, WP3.3, WP3.4, WP3.5, WP3.6, WP3.7
Dissemination Level:	PU = Public
Planned Delivery Date:	M12/30 September 2014
Actual Delivery Date:	M15/15 December 2014

Authors:	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), Satu PALVA, UH (P86), Viktor JIRSA, AMU (P104), Mariano SIGMAN, CEA (P9), 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), Lars NYBERG, UMU (P56), Johan ERIKSSON, UMU (P56), Neil BURGESS, UCL (P71), Fabian CHERSI, UCL (P71), Yves FREGNAC, CNRS (P7), Brice BATHELLIER, CNRS (P7), Christophe PALLIER, CEA (P9), Riitta HARI, AALTO (P2), Lauri PARKKONEN, AALTO (P2)
STO Review:	UHEI (P45): Björn KINDLER, Sabine SCHNEIDER, Martina SCHMALHOLZ
Editorial Review:	EPFL (P1): Richard WALKER, Guy WILLIS, Celia LUTERBACHER

Abstract:	This report is the Month 12 Deliverable for the HBP Subproject 3, Cognitive Architectures. The Deliverable, entitled “Implementation and validation of functional mapping methods” describes SP3’s main achievements and plans for the next six months. Based on this report, SP3 will make key contributions to a successful operational phase of the HBP.
Keywords:	Cognitive architectures, Milestones, Key Performance Indicators, strategic experimental protocols

Document Status

Version	Date	Status	Comments
1.0	26 Sep 2014	Draft	Report completed by Thiên-Ly Pham and reviewed by Stanislas Dehaene (SP Leader)
2.0	8 Oct 2014	Draft	Reviewed by Guy Willis and STO
3.0	24 Nov 2014	Draft	Report reviewed by Thiên-Ly Pham
FINAL	4 Dec 2014	Final	Final check by Richard Walker & Guy Willis

Table of Contents

List of Figures and Tables

1. Introduction

1.1 The Human Brain Project (HBP)

The Human Brain Project (HBP) is a major international scientific research project, involving over 100 academic and corporate entities in more than 20 countries. Funded by the European Commission (EC), the ten-year, EUR 1 billion Project was launched in 2013 with the goal "to build a completely new ICT infrastructure for neuroscience, and for brain-related research in medicine and computing, catalysing a global collaborative effort to understand the human brain and its diseases and ultimately to emulate its computational capabilities."

The fields of neuroscience, medicine and information technology each have important roles to play in addressing this challenge, but the knowledge and data that each is generating have been very fragmented. The HBP is driving integration of these different contributions.

During the Ramp-Up Phase, the HBP will collect strategic data, develop theoretical frameworks, and perform technical work necessary for the development of six Information and Communication Technology (ICT) Platforms during the Operational Phase. The ICT Platforms, offering services to neuroscientists, clinical researchers and technology developers, comprise Neuroinformatics (a data repository, including brain atlases and analysing tools); Brain Simulation (building ICT models and multi-scale simulations of brains and brain components); Medical Informatics (bringing together information on brain diseases); Neuromorphic Computing (ICT that mimics the functioning of the brain); and Neurorobotics (allowing testing of brain models and simulations in virtual environments). A High Performance Computing Platform will support these Platforms.

1.2 HBP Subproject 3: Cognitive Architectures

SP3 Cognitive Architectures belongs to the Data pillar of the HBP. The goal of SP3 is to select well defined, challenging cognitive domains already partially studied by cognitive neuroscience, and to refine the understanding of their “cognitive architecture” (areas, circuits, internal codes, dynamics). To achieve this aim, the scientists in SP3, each specializing in a specific cognitive domain, have two main goals:

- Review the existing literature on the specified cognitive function
- Define new strategic experimental protocols (previously called “localisers”) to dissect the associated patterns of brain activation and response dynamics. The observed patterns of activation and dynamics should make it possible to identify a) the brain regions involved in the task, b) the likely circuitry connecting these brain regions, and c) the principles of information processing within and between these brain regions.

We refer to this information collectively as the cognitive architecture for the task. By assembling all the top-down constraints that arise from our knowledge of behaviour and brain circuits, cognitive neuroscientists in the HBP help create and constrain theoretical models, possibly framed in the form of computer simulations that capture and reproduce the main facts about a cognitive architecture.

SP3 interacts primarily with two other Subprojects in HBP:

- SP3 provides strategic experimental protocols to SP2, which will select a subset of them to be run by the Individual Brain Charting task, i.e. a fixed set of human subjects imaged at NeuroSpin (Bertrand Thirion; Task 2.1.1).

- SP3 provides constraints for the cognitive brain models developed by SP4, and three actual simulation models (for action recognition, spatial navigation, and multisensory integration). Models of cognitive architectures will span scales ranging from high-level conceptual models to more explicit models with individual simplified neurons.

1.3 Purpose of this Document

The report will provide a detailed account of the implementation and validation of the methods described in D3.7.2 Implementation and validation of functional mapping methods.

1.4 Structure of this Document

The remainder of this chapter provides an SP-level overview, highlighting the SP's main accomplishments and issues encountered in the period M1-M12. Subsequent chapters look at accomplishments and issues within individual components of the SP, as defined in D3.7.1 Methods, Indicators of Progress and Target Values for Functional Mapping of the Human Brain and Derivation of Cognitive Architectures Selected for Cognitive Functions.

SP3 consists of 7 Work Packages (WP):

- WP3.1: Perception-Action
- WP3.2: Motivation, Decision and Reward
- WP3.3: Learning and Memory
- WP3.4: Space, Time and Numbers
- WP3.5: From sensory processing to multimodal perception
- WP3.6: Capabilities characteristics of the human brain
- WP3.7: Scientific coordination

The Annexes present in tabular form what the Subproject planned to achieve in this period and what it actually achieved, including the Subproject's Scientific Key Performance Indicators (S-KPIs).

1.5 Overview of Subproject 3 Achievements

Stanislas Dehaene (CEA), SP3 leader and Thiên-Ly Pham (project manager, CEA) coordinated SP3 activities. They ensured that the Subproject remained on track up to Month 12, especially with regard to its principle activities: the creation of draft cognitive architectures and strategic experimental protocols.

To maximize the scientific utility of these documents and ensure their peer review, Stanislas Dehaene proposed that they should be published together as a special issue of a major journal. The whole, consisting of about 15 reviews, will be a major contribution to cognitive neuroscience. Stanislas Dehaene and Yadin Dudai (WIS) will be the editors of this special issue. Detailed discussions have been started with the editor of the journal *Neuron*, Katja Brose, who strongly supports SP3's desire to publish its work.

A second Milestone for SP3 is the delivery in Month 12 of strategic experimental paradigms ("localizers") for each of the major cognitive functions under study. After reviewing the literature and, in many cases, running new experiments, the Task leaders have provided descriptions of their best fMRI or MEG experiments, designed to parse their cognitive

function of interest. The scientific coordinator, Stanislas Dehaene, will transmit these documents to Bertrand Thirion (Inria), who is in charge of the Individual Brain Charting project in SP2.

1.5.1 WP3.1 Perception – action

Task 3.1.1 involves three separate groups: Malach's (WIS), Dehaene's (CEA) and Fries' (ESI). Using an on-going auditory detection task in conjunction with naturalistic visual stimulation, the Malach group was able to use these patterns to uncover putative long-term predictions in the human visual cortex. Thus, we have devised and demonstrated the utility of a powerful new experimental approach for uncovering long term predictions in the visual cortex. This experimental paradigm could be an important part in the planned "localizer" toolbox planned for the HBP project.

In Dehaene's group, a MEEG and psychophysic experiment was set up to investigate the invariance property of object recognition and to test Bayesian models of visual perception. The behavioural version of the paradigm has been developed and psychophysic data were collected. A second pilot of the MEEG version of the experiment was conducted. The scripts for pre-processing and analysis (ERF/ERP and decoding methods) have been developed and preliminary results are expected in the near future. For the next six months, the data acquisition of this experiment and their analysis will be completed.

Pascal Fries' lab has investigated the role of oscillatory synchronisation in inter-areal coupling. We have submitted a manuscript, 'Stimulus induced visual cortical networks are recapitulated by spontaneous local and inter-areal oscillatory synchronisation', which finds that interactions between early and mid-level visual areas are constrained to specific frequency bands both during stimulation, as well as during passive fixation. This finding supports our hypothesis that distinct frequency bands are involved in the coupling of distributed neural populations. In order to further test specific predictions of the revised CTC hypothesis, we have collected preliminary data from anesthetised cats that allow analysis of the laminar distribution of inter-areal coupling, as well as the modulation of specific oscillations via optogenetic intervention. These new data will allow us to further test and constrain models of inter-areal interaction. We have achieved our Milestone of defining strategic data for the evaluation of models of inter-areal information transfer and routing and have proceeded in pilot studies to acquire the needed data.

In T3.1.2, the Giese group (EKUT) has substantially extended a physiologically inspired model for the visual recognition of actions and the perception of causality from abstract stimuli. A new mathematical framework based on multi-dimensional neural fields has been introduced for the modelling of multi-stability in action perception. As further components, noise and different forms of adaptation (firing-rate fatigue and input fatigue) have been introduced to model exactly neurophysiological data from relevant areas in macaque cortex (Note: acquisition of non-human primate data is not funded by HBP). A new framework for the analysis of the stability of solutions in two-dimensional neural fields, which is based on a level-set approach, has been adopted for the analysis of bifurcations of the network dynamics of the model, resulting in switches between different travelling pulse solutions. A new perceptual illusion has been discovered, showing that action perception is dependent on a 'lighting-from-above prior', and a neural model has been developed that accounts for the illusion, and which adds a shading pathway to a previous neural model for body motion perception.

In T3.1.3, the Blanke group (EPFL) achieved two main goals over the last year. Firstly, it performed an anatomical classification of the right and left temporo-parietal junctions (rTPJ and lTPJ) that have previously been associated with the bodily self (self-location and first-person perspective; Ionta et al., 2011). Secondly, it thoroughly reviewed strategic data about anatomical and functional constraints of multisensory integration in non-human

primates. Over the next six months, it will extend the anatomical dissection of the TPJ to the functional domain, and perform a meta-analysis of coactivations with related cognitive functions. The Slater group (UB) has carried out an experiment to explore the implications of embodying participants into avatars, and explore aspects of brain activity (surface EEG signals) when, in the context of virtual body ownership, a virtual arm moves while the corresponding real arm is static. The objective is to understand the relationship between real, and embodied virtual arm movements. A secondary goal is to examine the extent to which this embodiment increases the activity of the internal imitation process, i.e. mirror neurons, with important outcomes in the field of rehabilitation. Data was collected on 18 male subjects. The group analysed this with standard ERP analysis, but it is also applying methods of 'Big Data' analysis using genetic programming. With this massive amount of data, it is contributing to knowledge about how the brain represents the body, and in particular to the field of illusory agency.

Following the competitive call, new Partners (SLIMM and SpaceTime Projects) joined WP3.1 in April 2014.

The goal of the SLIMMM project (T3.1.4) is to make empirically informed neural network models of visual cortex, assess the output of this model as a function of input variations, and compare network output with detailed recordings from neural networks in monkeys (Note: acquisition of non-human primate data is not funded by HBP) and with recordings of population responses from humans (MEG). From these comparisons, De Weerd (UM), Roberts (UM) and Hadjipapas (UNIC) groups aim to derive which detailed network properties are likely implemented in human cortex and could underlie the human MEG response. The group is currently focusing on building a single-layer and spatially non-differentiated PING model using realistic neural firing frequencies and firing patterns for E and I model-neurons. This is an important aspect of making (P)ING networks empirically valid, which is currently insufficiently addressed in the field. We are currently optimising approaches to automatically classify different neurons (recorded on depth probes in NHP V1) into E and I categories. At the same time, the group is already doing specific modelling experiments/simulations to uncover the principles by which network synchronisation occurs, and specifically, what synchronisation plays in organising E and I population activity.

During the first six months of the project (April 2014-Sept 2014, with funding from June 2014), all KPIs in the SpaceTime Task (T3.1.5) have been met. Palva (UH) and Jirsa (AMU) groups have defined the experimental paradigm and generated the threshold stimuli. To validate the task, we acquired psychophysical test-retest reliability data with good results and a manuscript in preparation. The manuscript describing the SEEG method to-be-used in HBP is provisionally accepted. The group also completed initial analyses of resting-state SEEG data (N=22), prepared two manuscripts, and begun pre-processing resting-state data from new SEEG subjects (N=16). Reduced network models have been formalised and computationally implemented. First network simulations with time delays were performed with good results confirming initial hypotheses.

1.5.2 WP3.2 Motivation, decision and reward

T3.2.1 has worked on understanding how we compute confidence in simple perceptual decisions. The focus of the work during this year, in collaboration with the Mainen and Costa groups (FCHAMP) and Dehaene and Sigman groups (CEA), has been on designing tasks that satisfy three general requirements: 1) they can be performed in humans and non-human animals, to investigate confidence in different models, 2) they probe the main computational mechanisms and serve to decide between different models of decisions confidence and 3) they rely on stimulus parameters which can be mapped reliably in the

fMRI, so that this project can synergise with large scale localisation of functions in the human brain.

Following the Competitive Call, WP3.2 welcomed as new Partners the nCDN and SELF-MODE Projects.

T3.2.3 goal is to determine how phasic modulatory effects from the brainstem locus coeruleus (LC) noradrenaline (NA) system shape decision processes in the cerebral cortex. During the first six months, the Donner (UVA) and Engel (UKE) groups have collected complete high-resolution fMRI data sets covering the LC and large parts of the cortex, entailing three elementary decision-making tasks. The team has successfully implemented a brainstem-specific preprocessing pipeline (incl. physiological artefact removal, and brainstem realignment), sorted trials into high vs. low pupil dilation (a proxy of LC-activity), and delineated the LC and several cortical areas of interest in each individual brain. The Task members also collected a combined pupillometry and whole-brain MEG data set using a classical (random dot motion) decision-making task. Finally, the team began to simulate the effects of neuromodulation on cortical decision computations by extension of the leaky competing accumulator (LCA) model. This model accurately reproduces the empirically observed relationship between pupil dilation and choice behaviour.

In T3.2.4, the Hendler group (TASMC) has the task to map motivational decision-making process at a multi-level approach: from single cell to fMRI. During this half-year, Task members prepared the protocols for data acquisition from epileptic patients. They have adjusted their psychological paradigms to fit the multiple level recordings. Finally, they began to record data from patients implanted with depth electrodes. The data from one grid patient and two depth electrode patients have been collected and initial data analysis has been performed. These initial results have been presented at the HBP educational meeting in Tel Aviv, June 2014 and in the HBP annual summit at Heidelberg. Progress is according to plan.

1.5.3 WP3.3 Learning and Memory

In T3.3.1, the Karni group (UHAIFA) has tested the conjecture that the modulation of activity in a given brain area (M1) by task repetition reflects learning and overnight procedural memory consolidation (Experiment 1). The reanalysis of past and new data from imaging experiments shows that brief but robust modulations of M1 activity (Gabitov *et al*, JCN, 2014) 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 stabilisation of activity upon sequence repetition reflects short-term familiarity with novel movement sequences (i.e., the movement syntax). A paper is in press (Gabitov, Manor, Karni). Experiment 2 addresses motor cortex plasticity driven by visual input (action observation). The behavioural data suggest that executing and observing movements improve task performance and trigger skill consolidation processes. However, consolidation can be blocked by ensuing action but not by observation, indicating that skills acquired in doing or observing do not overlap in the brain (Maaravi-Hesseg, Gal, Karni).

In T3.3.2, the group of Yadin Dudai (WIS) developed the first paradigm to separate computational modes of hippocampal function in human episodic experience, which probably reflect pattern separation/completion. This is essential for simulating brain circuits of memory formation. An fMRI localiser for engagement of distinct hippocampal sub-regions in these mnemonic functions was developed. The group of Rony Paz (WIS) identified how memory generalises differently under positive vs. negative contexts. The group of Jan Born (EKUT) assembled the first model of the generation of “spindle-ripple

events", which mediate transfer of reactivated memory from hippocampus to extrahippocampal circuitry in consolidation during sleep.

In T3.3.3, Nyberg and Eriksson (UMU) have submitted a first draft of the cognitive architecture and localiser protocol for working memory. While the localiser protocol remains unchanged, they have shifted the focus of the locally performed fMRI experiment. Specifically, the team focused less on working-memory capacity limitation and instead explore the possibility of non-conscious working memory. For this, data collection and initial analyses has been completed and we are thus not behind schedule despite the change. The team has not experienced any major problems. To finalise Deliverables and Milestones in time, T3.3.3 has started working earlier than stated in the budget plan.

1.5.4 WP3.4 Space, time and numbers

During the first year of the HBP project, the Burgess group (UCL) attempted to define the cognitive architecture of spatial navigation, by reviewing the relevant literation. It characterised the relevant brain systems, representations and learning rules, and identified an appropriate 'functional localiser' task for neuroimaging, consisting of object-location memory in virtual reality. The group began to develop a neural-level computational model of spatial navigation, implementing the above cognitive constraints. We aim to validate the simulation by fitting behavioural data on mammals learning spatial locations in various situations, while retaining plausibility at the simplest neuronal level. This work combines the interests of SP3 and SP4. The modelling work will be in part suitable for implementation on the large-scale network models developed in SP6 in the neuromorphic computing devices implemented in SP9. The group has encountered no particular difficulties so far.

1.5.5 WP3.5 From sensory processing to multimodal perception

In T3.5.1, the data obtained earlier with intracellular techniques have been fully processed and published. The intracellular data have been used to build a comprehensive model of aspects in higher mammals (cat, ferret, monkey) that are applicable to man. A parametric search data-driven model of V1 has been finalised to account for the receptive field dynamics of first-order simple cells receiving direct input from the thalamus. T.3.5.2 used two-photon calcium imaging to record population activity in visual cortex during auditory and visual stimulation. The data shows that numerous mouse V1 neurons respond not only to visual but also to auditory stimuli and that bimodal stimulation often leads to non-additive (non-linear) responses.

1.5.6 WP3.6 Capabilities characteristic of the human brain

In T3.6.2, Pallier and Meyniel (CEA) drafted the content of a review on the neuronal implementation of different aspects of sequence processing, ranging from the extraction of transition probabilities between stimuli, to nested structures. The group also achieved significant progress in its experimental work. Experiment 1 examines with fMRI how human extract the regularities of linguistic nested-structures, experiments 2 and 3 examine how humans extract the regularities of non-linguistic structure. Experiment 2 examines this capability with MEG data, using sequences structured by the transition probabilities between the stimuli. Experiment 3 investigates this capability with behavioural and fMRI data, using sequences structured by chunks and nested rules.

Task 3.6.3 addresses the importance of other people in shaping our mind. The Task started officially on Month 9 of the project. The group of Riitta Hari (AALTO) has started to develop an fMRI localiser for social brain functions using simplified social stimuli such as movie clips of persons and animated characters. The group has also refined its two-person measurement set-ups for studying interacting subjects in MEG and fMRI, collected pilot

data, and developed analysis methods for these dual recordings. Collaboration with prof. Martin Giese, leading Task 3.1.2, has been started.

1.6 Overview of Subproject 3 Problems

SP3 did not encounter any problems during this first year.

1.7 The Next Six Months for Subproject 3

At Month 18, SP3 will provide detailed cognitive architectures for each of the Tasks studied in the Subproject together with simplified models, to be developed in tight coordination with SP4.

Functional mapping data, cognitive architectures and models for the HBP Human Brain Atlas, which will be deposited in the HBP Human Brain Atlas, will include preliminary data on the cognitive architectures for each of the Tasks studied in the Subproject together with simplified models of spatial navigation, multimodal integration, and visual action recognition, to be developed in tight coordination with SP4.

2. Perception-Action (WP3.1)

2.1 Study of the Circuits Involved in Non-Conscious and Conscious Mechanisms of Visual Recognition (T3.1.1)

This Task involves the following 2 sub-Tasks, each of which is reported on separately:

- Visual Perception
- Visual attention and the mechanisms of inter-areal communication

2.1.1 Visual Perception

Mapping long and short-term perceptual predictions of the human visual cortex- a combined MEG and fMRI approach (The Malach and Dehaene groups).

2.1.1.1 Research goal

As stated, the overall research goal of this Task is to achieve a detailed map of the circuits involved in both long and short-term predictions in visual recognition. It is becoming evident that a critical aspect of visual perception is the implementation of *a-priori*, predictive information in neuronal circuits. Such prior information, which is postulated to be implemented in the form of "internal models" of objects and is likely to be operating below the threshold of awareness, allows the visual perceptual system to operate very efficiently and adaptively, behaving similarly to an optimal Bayesian inference device. This capability is particularly important when faced with ambiguous or noisy information. However, the specific details and neuronal implementation of this predictive information processing remains to be elucidated. In the first year of the HBP project, we have made major advances towards achieving this ambitious goal. These include the development of experimental paradigms that will allow the uncovering such predictive information in the functional responses and neuronal signals of the visual system. Importantly, new developments in the field, by HBP Partners and others, have demonstrated new possibilities for uncovering long-term visual prediction mechanisms through brain imaging.

2.1.1.2 Main achievements (SP3_SKPI-01)

Malach's group (WIS) has achieved all the targeted Milestones set by the HBP and even exceeded them. We, along with a number of other groups, have shown that studying the patterns of spontaneously emerging connections in the human cortex actually constitutes a powerful tool for mapping long-term cortical prediction mechanisms, as embedded in synaptic connectivity[1]. We were able to use this powerful new "window" to examine the internal predictive models of human visual representation, using a specially designed experimental paradigm (Figure 1), based on an on-going auditory detection Task, in conjunction with naturalistic visual stimulation. We were able to use these patterns to uncover putative long-term predictions in the human visual cortex. The elegant aspect of applying this approach to the visual system is that when mapped onto the orderly retinotopic representation of the visual cortex, such predictions can be literally projected out to visuo-topic space and can thus be deciphered intuitively. In summary, we have devised a powerful new experimental approach for uncovering long-term predictions in the visual cortex and demonstrated its utility. This experimental paradigm could be an important part in the HBP's planned "localiser" toolbox.

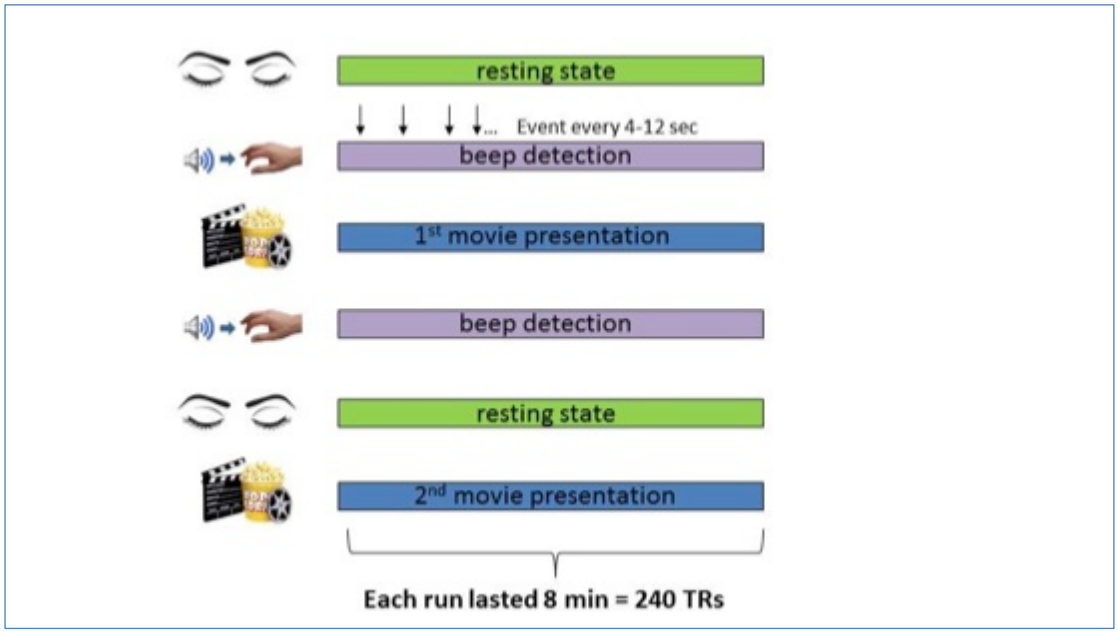


Figure 1: Uncovering long-term visual biases and predictions in the human brain

A schematic illustration of the experimental brain-imaging paradigm proposed to uncover long-term visual biases and predictions in the human visual cortex. The experiment consists of six scans. Two scans are aimed at simulating the coherent activation patterns that occur during daily naturalistic vision, which are conjectured to set up long-term predictions (reflected in coherent activations) of the human visual system. These patterns can be detected by mapping recurrent correlation patterns that are correlated across repeated presentations of an identical movie segment (blue bars). Resting state scans (green bars) are aimed at mapping spontaneously emerging patterns that are conjectured to recapitulate such long-term predictions. Finally, the "beep detection" scans are aimed at revealing such spontaneous visual biases under cognitively controlled conditions, which, due to the different (auditory) modality, should not interfere with the spontaneous emergence of visual biases. Preliminary results demonstrate that this paradigm is highly effective at uncovering internal biases emerging both during naturalistic vision and spontaneous visual cortex fluctuations.

In Dehaene's group (CEA), a MEEG and psychophysical experiment was set up to investigate the availability of internal models of objects and their role in the invariance property of object recognition. The experiment recorded MEG signals (and possibly fMRI and ECOG signals in the future) while participants examined an item in slow rotation on screen (either a face or an object, rotating clockwise or counter-clockwise at 12 rpm). We examined two time scales of prediction: long time scales, reflecting the internalisation of natural world statistics of objects such as faces, and short time scales, reflecting "on-line" acquisition of new expectations about a specific object.

To study long-term predictions, we employed natural, dynamic changes in an object, e.g. a rotating head, whose geometry is familiar to the viewer. To study short-term predictions, the paradigm also requires learning one new view of the object (the front and profile views of the object conform to everyday life, but subjects also learn that the back view comprises an unusual pattern). The different views of two objects (a face and a coffee machine) are learned during an initial exposure period. The data acquired during this period of the experiment should allow us to train multivariate decoders for each object and each orientation, and using generalisation of decoding, to test for the presence of an invariant representation of the object, regardless of which particular view is presented (familiar or recently acquired). During the main experiment, while the object is rotating, an occluder hid the object for a variable duration (¼, ½, ¾ or a full turn). Using multivariate decoding, this period allowed us to probe the presence of an internal model of the rotating object: can we continuously decode what object is occluded and what is its orientation? Furthermore, when the screen drops, it reveals either the appropriate object

or the wrong object, in either the appropriate rotation or the counter-rotation, and with either the appropriate angle or a different angle. If the predictive coding hypothesis is correct, each of those mismatches should elicit error signals detectable via MEG.

The behavioural version of the paradigm has been developed and psychophysical data were collected. Pilot data were collected using the MEEG version of the experiment. The scripts for pre-processing and analysis (ERF/ERP and decoding methods) have been developed and preliminary results indicate that a considerable amount of object decoding can be achieved from MEG signals. In the near future, we will obtain results on the decoding of the internal model during occlusion.

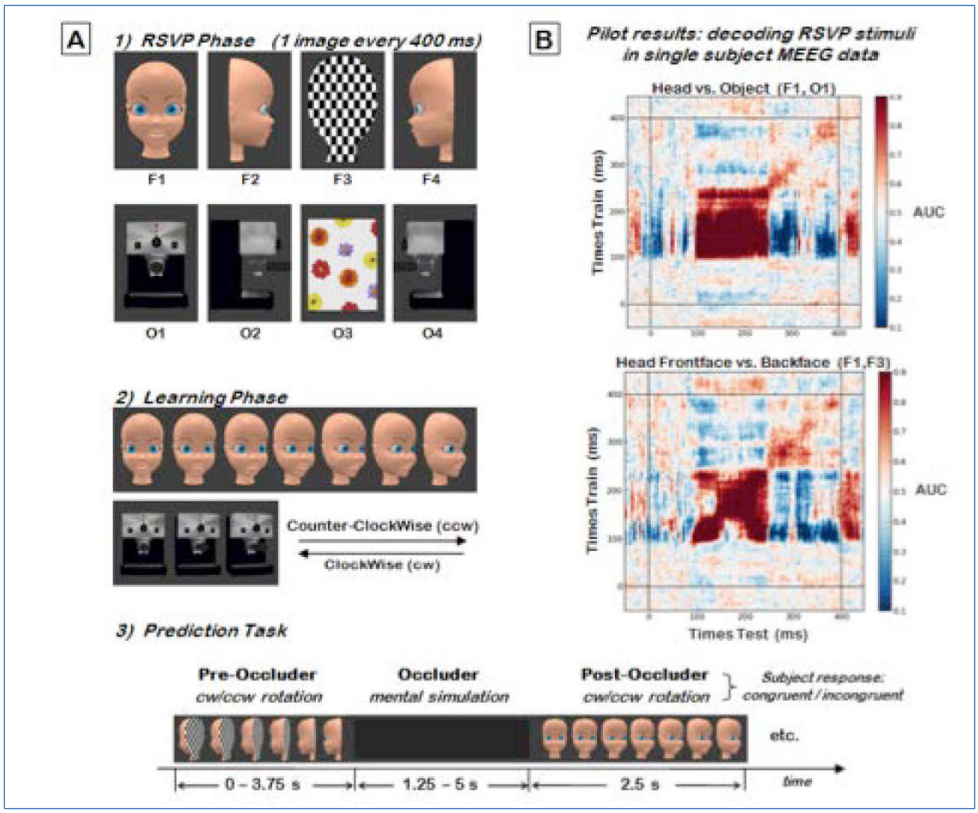


Figure 2: Visual prediction: experiment design and analysis tool

A) Experiment Design. 1) RSVP Phase. Rapid Serial Visual Presentation (RSVP) enables the acquisition of 60 event-related fields/potentials for all the four cardinal views of each object (F1, F2, F3 and F4 for the head; O1, O2, O3, O4 for the coffee machine), one image being presented every 400 ms. These data will be used as spatio-temporal localisers. *2) Learning Phase*. The participant is introduced to the continuous stimuli within four sequences: one for each object and rotation direction (clockwise - cw - and counterclockwise - ccw). 3) Predictive Task. While the rotating stimuli are occluded, the participant is asked to simulate mentally the continuation of the rotation. When the stimuli reappear, the subject has to response concerning the congruency (the stimulus viewpoint is in the configuration expected at the time of the occluder disappearance) or the incongruency (wrong viewpoint, rotation direction or object). The numerous independent variables are balanced and crossed. B) Analysis tool. The stimuli are classified with Support Vectors Machine decoders for each time of the trials and also applied to all the other time slots (called generalisation across time). The classification of the objects, and of the views of a same object, is tested. The high efficiency of the decoders (Area Under the Curve - AUC - > 0.95) in the two conditions and its stability across time suggest that decoding the objects and their views during both visible and mental simulation rotations is possible, as a first step of additional analysis.

2.1.1.3 Plans for the next six months

During the next six months, MEG data acquisition and analysis will be completed. Decoders will be built to distinguish from brain activity which object is present behind the occluder,

and at what angle it is presented. Using the technique of generalisation of temporal decoding[2], we will decipher the temporal unfolding of these internal coding signals and determine how quickly they are revised depending on the (detected or undetected) sensory error. The results will provide strong constraints on the localisation and timing of both internal representation and error detection mechanisms in invariant visual perception. Future work will examine how to generalise these results to the perception of non-conscious (subliminal) views of the items.

2.1.2 Visual attention and the mechanisms of inter-areal communication

2.1.2.1 Research goal

In order to behave adaptively in the world, animals are forced to distinguish between relevant and irrelevant sensory information and to select the relevant aspects for further processing. The implementation of this selection process and the reliable routing of selected information require the coordination of neuronal activity in both a top-down (for selection) and bottom-up (for routeing) manner across many distributed populations. A prominent hypothesis for selective inter-areal interactions is communication-through-coherence (CTC), which proposes that information exchange between neural populations is regulated by the degree to which the populations are rhythmically synchronised[3]. In vision, attention is thought to select relevant features, locations or objects and to selectively route information from the primary visual cortex to higher visual areas in order to facilitate processing. In the case of visual-spatial attention, it has been demonstrated that synchronisation between early and late visual areas occurs selectively between neural populations representing the attended visual stimulus[4,5]. However, a mechanistic account of both the top-down and bottom-up components of this process are lacking. We are examining the patterns of local and distributed dynamics that give rise coordinated information routing and selection.

2.1.2.2 Main achievements

The laboratory headed by Pascal Fries (ESI) is investigating the role of oscillatory synchronisation in inter-areal coupling. We have submitted a manuscript, '*Stimulus induced visual cortical networks are recapitulated by spontaneous local and inter-areal oscillatory synchronisation'*, which finds that interactions between early and mid-level visual areas are constrained to specific frequency bands, both during stimulation, as well as during passive fixation[6]. This finding supports our hypothesis that distinct frequency bands are involved in the coupling of distributed neural populations. We have reviewed current evidence for inter-areal coupling, specifically, at non-zero phase lags and this has resulted in a modification of the CTC hypothesis for distinct feed-forward and feedback streams[7]. In order to further test specific predictions of the revised CTC hypothesis, we have collected preliminary data from anesthetised cats that allow analysis of the laminar distribution of inter-areal coupling, as well as the modulation of specific oscillations via optogenetic intervention. These new data will allow us to further test and constrain models of inter-areal interaction. We have achieved our Milestone of defining strategic data for the evaluation of models of inter-areal information transfer and routing and have proceeded in pilot studies to acquire the needed data.

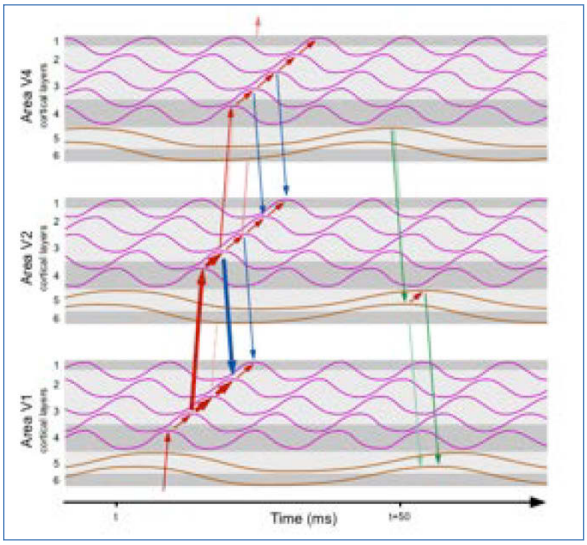


Figure 3: CTC with inter-areal delays between hierarchically arranged areas[7]

Schematic illustration of the modified CTC hypothesis incorporating delays between areas and between the layers of a given area. Purple lines illustrate supragranular gamma oscillations with a systematic delay from layer 4 toward layer 1. Red arrows indicate feedforward signals, entering in the lower left into layer 4, and proceeding through supragranular layers onwards to layer 4 of the next higher area. Blue arrows indicate supragranular feedback between closely neighboring areas. The thick red and blue arrows highlight one complete cycle of feedforward and reentrant feedback signalling. Note that the reentrant feedback arrives at the excitable phase of the local gamma, because it targets more superficial sub-layers, which are delayed relative to layer 4. Brown lines illustrate infragranular beta oscillations, green arrows infragranular feedback.

2.1.2.3 Plans for the next six months

Over the next six months, we will analyse our current data related to inter-areal information transfer and evaluate the degree to which explicit predictions from our current hypothesis conform to experimental findings. We will then begin to investigate specific mechanisms that may underlie the precise pattern of inter-areal dynamics we see. We will further perform additional experiments that specifically address the existence of consistent dynamic regimes at multiple spatial scales with targeted recordings with variable spatial resolution. These recordings will further allow us to constrain our model of inter-areal information flow by reconciling local micro-dynamics with the macro-dynamics observed across and between areas.

2.2 Understanding the circuits linking perceptions to actions (T3.1.2)

2.2.1 Research goals

The goal of Martin Giese's group (EKUT) is to refine and extend models of the visual perception of actions and, ultimately, to link such models to spiking network simulators developed by other members of the HBP. Furthermore, novel methods for the mathematical treatment of the underlying neural population dynamics have been established.

2.2.2 Main achievements (SP3_SKPI-02)

EKUT has substantially extended a physiologically-inspired model for the visual recognition of goal-directed actions and the perception of causality from abstract stimuli. A completely novel framework for the treatment of perceptual organisation phenomena in

action recognition has been established. The core of this framework is a two-dimensional neural field model (neural mass model) that represents body shape sequences together with the stimulus view. The resulting line attractor dynamics has multiple alternative time-dependent attractors, which correspond to different interpretations of ambiguous body motion stimuli[8]. We developed sophisticated models for the influence of intrinsic noise and adaptation on the switching dynamics. The modelling of adaptation was based on detailed modelling of electrophysiological data on adaptation effects in area IT[9]. Transferring such an adaptation mechanisms into a model for action recognition shows that perceptual switches in action recognition are likely driven by noise, not adaptation. In addition, the model accounts for the observed difficulty to find significant adaptation effects in action-selective neurons in area F5[10]. The strength of observed adaptation effects depends critically on the physiological basis of the adaptation (input fatigue vs. firing rate fatigue mechanisms. For the analysis of multi-stability in such dynamic neural representation we have established and refined a new level-set approach for the analysis of multi-dimensional neural field dynamics[10]. This approach was extended to fields with external input and travelling peak solutions, combining the original approach by Coombs et al. with a mathematical approach that links neural fields with stationary and traveling solutions[11]. A further substantial extension of a basic model accounts for the processing of shading cues in action stimuli. We added an additional visual pathway for the processing of inner illumination gradients within objects, suppressing the outer boundaries of the silhouette of the body. This model accounts specifically for a new visual illusion, which we confirmed experimentally, that shows that biological motion perception is critically dependent on shading cues and seems to be influenced by a lighting-from-above prior.

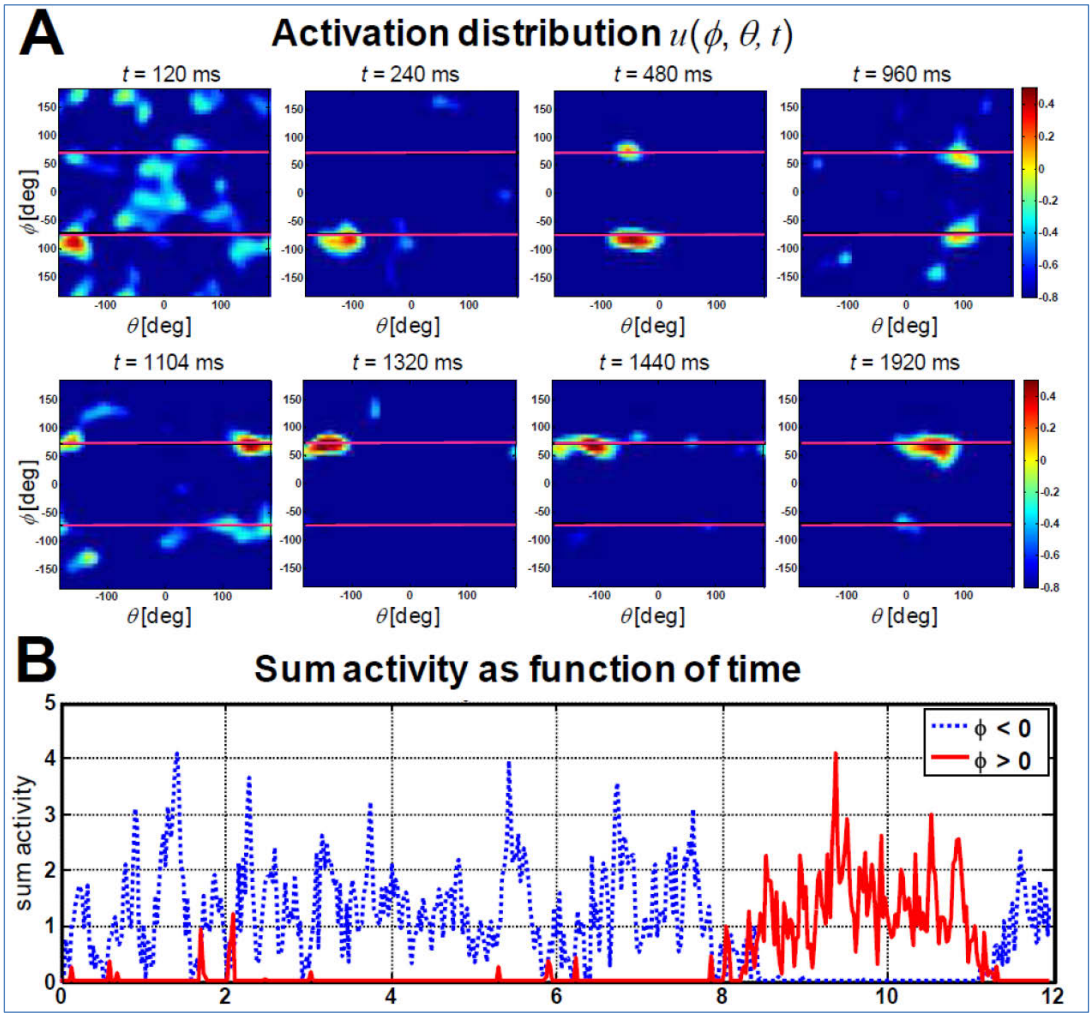


Figure 4: Two-dimensional neural field model reproducing multi-stability in visual action recognition[12]

A) Temporal evolution of the neural activity distribution in a neural field encoding snapshots of body shapes, where \theta corresponds to time within the gait cycle and \phi to the view angle. The model shows random switches between two stable travelling pulse solutions. B) Activity of the read-out neurons for the two alternative precepts. The read-out neurons sum over the part of the field representing positive respectively negative view angles. The read out neurons switch randomly between two stable perceptual states (perceived walking directions).

Due to the extremely small amount of funding we just now were able to hire a student for 18 months, who will be covered by other sources for the rest of his PhD.

2.2.3 Plans for the next six months

The model will be further worked out and the related psychophysical experiments will be finished and published. We have started to make contact with Markus Diesmann (Jülich) from the Theory Pillar, hoping to test some of his spiking network simulators to study spike-synchronisation effects in such dynamic action representations.

2.3 Understanding how Body Perception Becomes a Reference Point for the Sense of Self (T3.1.3)

2.3.1 Research goals

As stated in the DoW, this Task's aim is to refine our understanding of bodily self-consciousness by defining its constraints at the anatomo-functional level (SP3), relying on Bayesian models (SP4) and detailed mechanisms of multisensory integration (SP2; WP5). Experimental work and related applications will be performed in collaboration with the neurorobotics platform (SP10).

2.3.2 Main achievements (SP3_SKPI-03)

Olaf Blanke's group (EPFL) achieved two main goals over the last year. Firstly, we performed an anatomical classification of the right and left temporo-parietal junctions (rTPJ and lTPJ) that have previously been associated with the bodily self (self-location and first-person perspective)[13]. The motivation behind this approach is to be able to refer to and exploit data from non-human primate electrophysiology about the homologous brain regions, to better understand what lower and higher-level functions they subserve and how and why these regions presumably contribute to constructing bodily self-consciousness. Based on cytoarchitectonically defined regions of the human parietal operculum (OP)[14] and of the inferior parietal cortex (IPC)[15], we found that 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 (Figure 5). OP1, which is presumably the human analogue of area S2 in non-human primates[16], is considered to be a "perceptive" area strongly interconnected with the IPC[16,17]and potentially associated with some of the more complex functions of the OP, such as perceptual learning, tactile working memory and stimulus discrimination[18–22]. 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[23]. Effort is still ongoing to collect and integrate anatomical and functional data from humans. Secondly, we thoroughly reviewed strategic data about anatomical and functional constraints of multisensory integration in non-human primates. The main result from this work is that multisensory stimuli in space are coded in body-centred reference frames by peri-personal space neurons[24,25], in addition to bimodal and trimodal neurons - located particularly in the ventral premotor cortex (area F4[26,27]; also termed Polisensory Zone, PZ[27]), the posterior parietal cortex (in the fundus of the intraparietal sulcus, in the ventral intraparietal area VIP and in area 7[24,28,29]) and in the putamen[30].

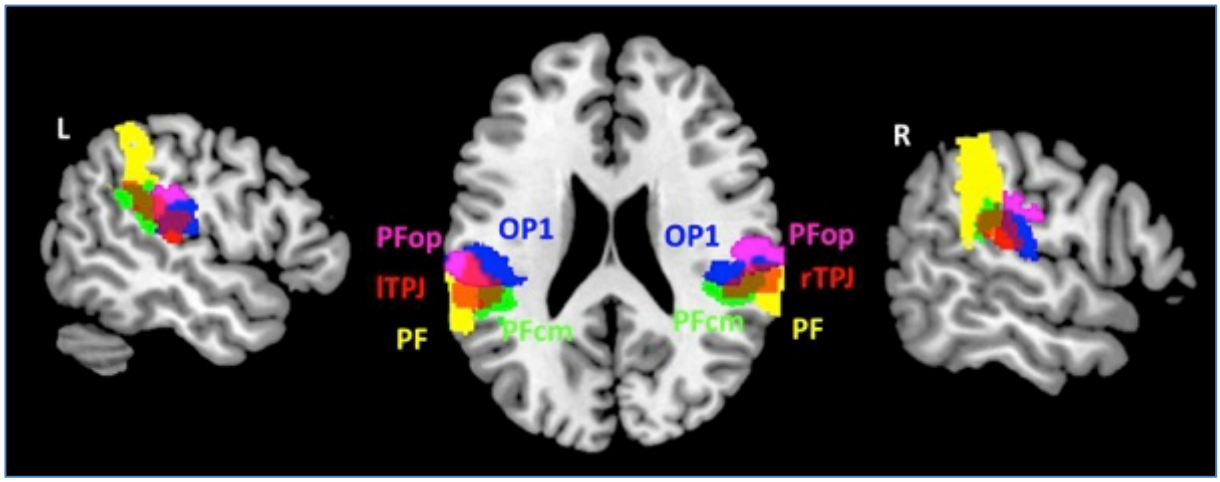


Figure 5: Anatomical overlap of regions encoding subjective sensation of self-location and first-person perspective (rTPJ & lTPJ), and cytoarchitectonically defined regions of the parietal operculum and the inferior parietal cortex

Mel Slater's group (UB) has carried out an experiment that utilising virtual reality (VR) aims to enhance previous results exploiting perspective as an important modulator of the effects of the mirror neuron system. Solid research has found that VR can be used to induce an illusion of ownership over a virtual body, making participants believe that they are in possession of a virtual body. We are particularly interested in what happens in conditions of strong body ownership, when the virtual body acts independently of the movement of the person's real body - for example, moving a limb, and whether there is a corresponding illusion of agency over the virtual body movements, and whether this is reflected in associated detectable brain activity. The objective is to understand the relationship between real, and embodied virtual arm movements. The experiment has 3 conditions: movement execution in a first person perspective view (1PP) over a virtual collocated body, movement observation in 1PP, and movement observation in third person perspective (3PP).

A major part of this is to explore methods of 'big data analysis' to analyse vast amounts of EEG data in an automatic way in order to discover equations that may shed light on the relationships between real and virtual body activity under different conditions. For this we use the Nutonian/Eureqa technology for genetic programming[1].

Data was collected on 18 male subjects. The setup of the experiment is shown in Figure 6. The study was approved by the ethics committee of Universitat de Barcelona. In the laboratory, participants were first placed in the same position as the avatar that they saw through the HMD that was co-located with their own body. The experiment consisted of 80 trials where either the avatar moved the hand or the participant moved the hand, i.e. where the participant was asked either to observe or to execute a movement. Movements lasted only 2 seconds, and within this time participants should have moved the arm towards the chin and back to the original position. The avatar movements for the observation condition were recorded with motion capture and also lasted 2 seconds. Each block of 80 trials targeted the same hand always, then after the 80 trials and a short pause the other hand was targeted for 80 more trials. This was particularly done to avoid error-monitoring systems, due to hand decision making, such as the Error Related Negativity.

[1] http://www.nutonian.com

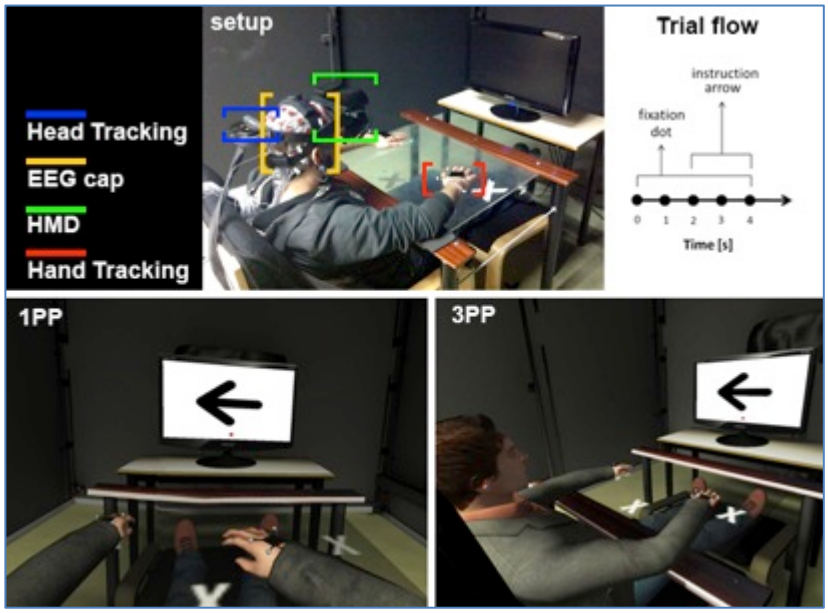


Figure 6: Virtual reality perspective experiment

On the top, we can observe one participant fitted with all the equipment. The participant was seating in the same position as the avatar (down). The two perspectives that were used for the experiment: the 1PP and the 3PP. On the top right, there is a small schema of the trial flow: participants fixated their view on the dot and performed the instruction that was showed in the screen. Every 10 trials, they had a specific time for blinking.

Participants went through 3 different conditions that were counterbalanced to avoid order effects.

1) Motor execution in Virtual Reality (VRME): participants entered the virtual environment and saw an avatar from 1PP, their task was to perform 80 times the motor action with one arm, and then with the other arm.
2) Motor observation in 1PP (MO1PP): participants entered the virtual environment and saw an avatar from 1PP, their task was to observe the motor action that was performed by the avatar, while not moving themselves.
3) Motor observation in 3PP (MO3PP): participants entered the virtual environment and saw an avatar from 3PP, their task was to observe 80 times the motor action that was performed by the avatar, while not moving themselves.

Finally an extra condition was run for a consistency check; in this condition the participant performed the motor execution in the real scenario without VR or HMD. For that condition, a real screen was used to indicate the trial flow to the participants (as shown in the setup -Figure 6). We will refer to this condition as real motor execution (RME). The whole experiment lasted for 2 hours in total, and the participants had the HMD removed between the conditions. One block of 80 trials lasted for approximately 13 minutes. Lateralised readiness potential's (LRP) were first studied as they have been shown to elicit activations as a result of the mirror neuron system[31,32] (Figure 7). A significant LRP was found for both motor execution conditions (RME and VRME) from the 300ms till the 600ms (t-test >0.05). No differences were found between the two conditions in the time course indicating that the LRP was equivalent in the real environment and in the virtual environment when the participant was executing the actions. Besides both the MO1PP and MO3PP elicit significant LRP at different timings (Figure 7). Of course, the voltage of the LRP during the observation of movements' conditions (MO1PP and MO3PP) is reduced when compared to the real execution conditions (RME and VRME). But interestingly, the first person perspective MO1PP has an earlier activation than third person observation MO3PP, and it is more aligned with the temporal dynamics observed in the case of the real motor

execution. This alone is a promising result towards the idea that virtual embodiment may be a significant factor describing the motor resonance mechanisms.

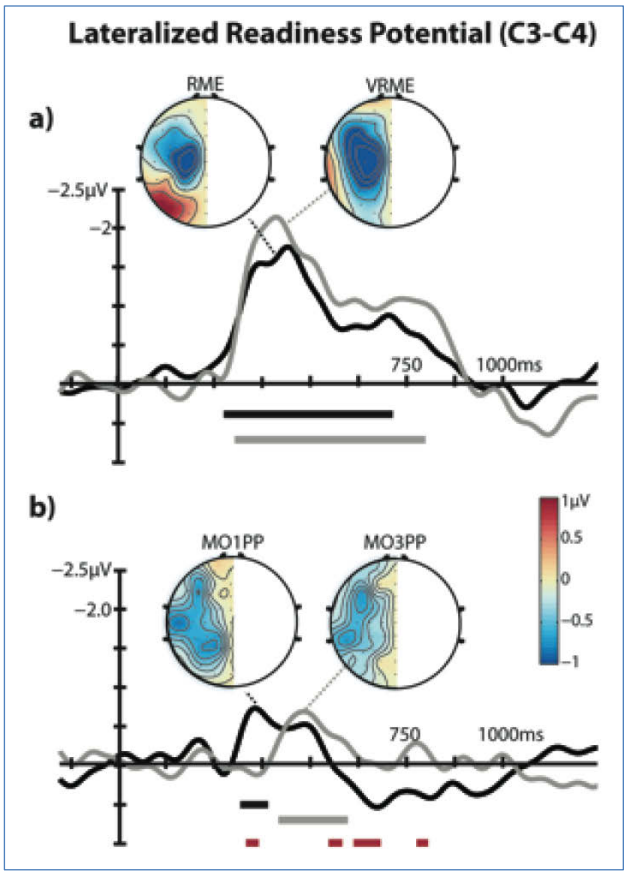


Figure 7: Great averaged lateralised readiness potential (C3-C4) for each of the conditions

A) RME (in black) and VRME (in grey) have a significant LRP from the 300ms till the 600ms (t-test >0.05). No differences were found between the two conditions in the time course. B) MO1PP (in black) and MO3PP (in grey) elicit significant LRP at different timings; MO1PP has an earlier activation than MO3PP. There were also differences between the two conditions (in red). Horizontal lines mark the times where significances were found>0.05.The topographies show the lateralization part projected on the left hemisphere, the graphs refer to the C3 electrode.

2.3.3 Preliminary Observations from the Genetic Programming Study

In order to become acquainted with this new method of analysis, we have applied the genetic programming technique (which is extremely computationally intensive and time-consuming) to the data set of one arbitrarily chosen individual. This consists of an 11480 x 64 data matrix of ERP values (64 electrodes). The time sequence is partitioned into the 8 segments as described above and reported in earlier Deliverables: (1) they moved their real right hand (outside of VR) (2) they moved their real left hand (outside of VR) (3) in the VR move right hand (4) in the VR move left hand (5) VR 1PP observed virtual right hand moved (real stationary) (6) VR 1PP observe left (real stationary) (7) VR 3PP observe right (real stationary) (8) VR 3PP observe left (real stationary).

For this preliminary analysis we have considered only periods 1, 5 and 7 - that is a comparison of the real hand moving, the real hand stationary but the virtual hand being seen to move by itself but from a first person perspective over the virtual body (with an illusion of body ownership), and the virtual hand moving by itself but where the body is seen from a third person perspective (i.e., there is no illusion of body ownership). Three

genetic programming exercises were therefore executed, one for each of conditions 1, 5 and 7. The number of generations of equations was at least 500,000 in each case, run on a 4-processor machine. Again our interest was to see if this works at all, and if so then much larger scale multiprocessor machines will be used to extend and deepen the search. In each case the genetic programming generated almost perfect fits to the data based on results from a very small number of electrodes: for (1) P07, O9 and Oz, for (5) Cz and C6 and (7) PO3, P7, C3, P5.

Even though the GP program was fed with many electrodes and ERPs for the different conditions some of the results can reflect previous findings on motor action observation. Such studies find that visual and kinaesthetic information may converge in parietal areas of the brain, and in our current GP analysis we find parietal electrodes to be significant descriptors of the mathematical equations, this is the case of PO7, P5, PO3 and P7. On the other hand motor cortex seems to also be contributing significantly to the equations also in the observation conditions as the role of electrodes Cz, C6 and C3 suggests.

Overall the method looks promising as a way of exploring these very large data sets, though at the time of writing we cannot draw conclusions about its longer-term viability.

2.3.4 Plans for the next six months

Over the next six months, Olaf Blanke's group (EPFL) will extend the anatomical dissection of the TPJ to the functional domain, and perform a meta-analysis of coactivations with related cognitive functions (perspective taking, visuo-spatial attention, empathy, theory-of-mind).

Mel Slater's group (UB) will continue to extend the data analysis and review this new method. Also the current results will be written up as a paper for publication. We will explore together with Olaf Blanke further studies on such illusory agency. If a simulated brain is to act in the world it has to have a body (whether virtual or robotic) and to the extent that it acts in the world in any way that mirrors animal action it must generate its equivalent simulated electrical activity in a way that mirrors what is observed in real animal and human studies. Hence we see this as a potential way in the long term of evaluating the output of the simulated brain.

2.4 Multi-scale data analysis and multi-scale transfer modelling (T3.1.4)

2.4.1 Research goals

The Task will use multi-scale data analysis and multi-scale transfer modelling to construct a spike/LFP-to-MEG transfer model. The Task will use this model to describe and validate human sensory circuitry constrained by LFP/spiking, and human MEG data obtained in the same sensory paradigm. The data generated by this Task will be uploaded to the neuroinformatics platform. This Task is led by Universiteit Maastricht and the University of Nicosia (Partners added via the Competitive Call).

2.4.2 Main achievements (SP3_SKPI-19)

2.4.2.1 Peter de Weerd's group (Universiteit Maastricht)

Background: When grating luminance contrast is increased there is a robust change in gamma oscillation frequency as measured by LFP in monkeys (depth probes and ECoG) and MEG in humans. In addition, a robust decay in gamma power is observed in the LFP[33] but not the MEG [34]. These phenomena are key to understanding the functional role of network frequencies and for investigating the stability of gamma oscillations at both local and

macroscopic levels. However, at the most basic level of spatially- undifferentiated neuronal models, it is not fully understood how excitatory (E) and inhibitory (I) neurons interact to generate network gamma oscillations. Moreover, it is unknown how increasing excitatory afferent drive modulates the interactions between E and I populations (as well as interactions within each population) to account for changes in frequency and power.

Aims: To understand the interactions between E and I populations, it is necessary to base the model on realistic spiking behaviour. To that aim, we are developing spike-sorting algorithms for our macaque extracellular recordings. Good algorithms to segregate neuronal populations into E and I classes will allow us to isolate two clusters of single neurons, namely regular spiking (putative excitatory) and fast-spiking (putative inhibitory) neurons. This is very important as it will permit the study of within –class (e.g. excitatory-excitatory) and between-class (excitatory-inhibitory) interactions and also class-specific coupling of the neurons to the population LFP.

Results: Various algorithms in the literature were implemented and tested such as the isolation of single units based on autocorrelation, clustering based on the up going and down going slope of the spike waveforms, and the peak-through distance [35]. Excellent isolation of single units could be achieved but some challenges ensued at the stage of clustering into distinct classes to be mapped on E and I cells.

Challenges: The classification strategies used in the literature did not lead to two classes that we could confidently map on putative E and I classes. To achieve this, we are now planning to introduce additional criteria, such as the relationship of unit spikes to the oscillation (spike- LFP phase coupling), the Fano factor, and the coupling between pairs of neurons (E-I) [36,37]. We are currently pursuing these avenues in both the empirical and the model data

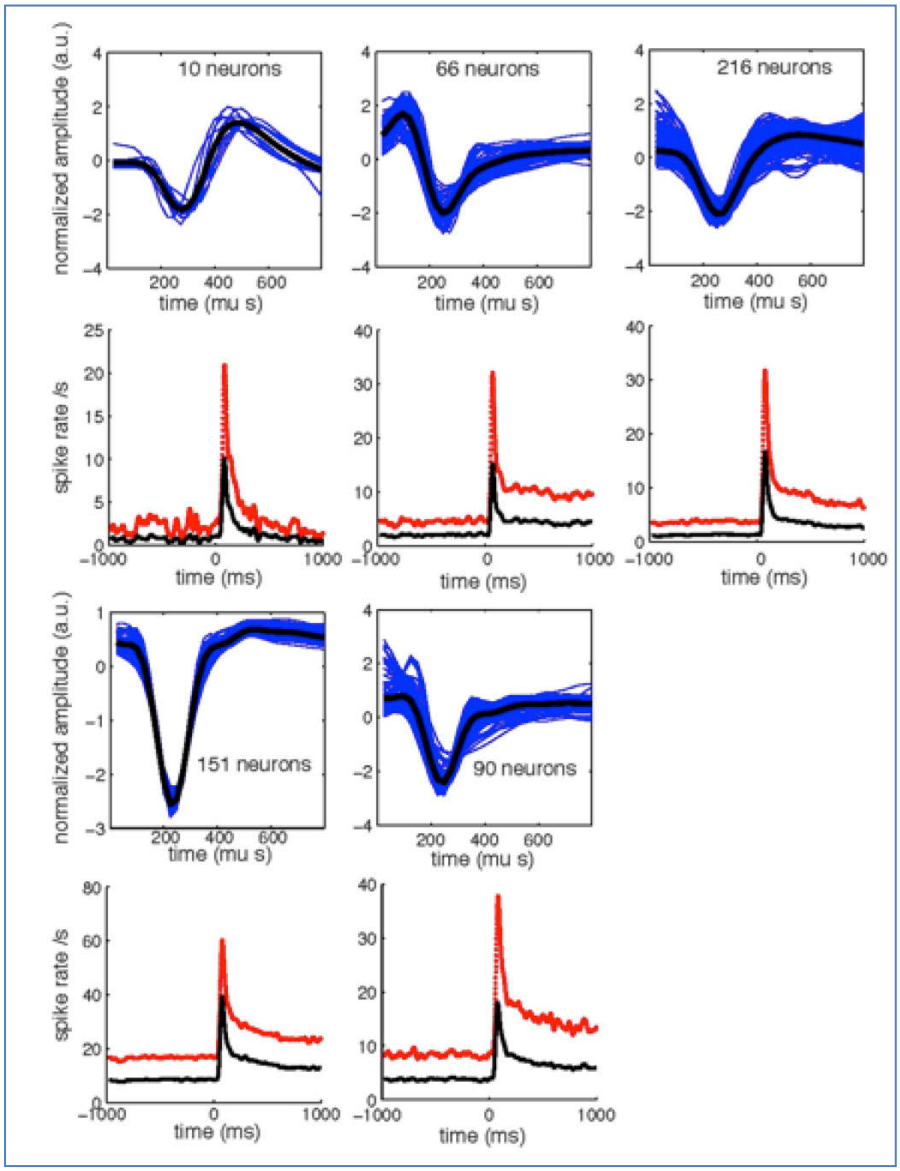


Figure 8: The average spike shapes (1st and 3rd rows) and firing rates (2nd and 4th rows) of five neuron types extracted from macaque visual cortex

In the top panels, the black line shows the average of spike shapes in each group and the blue lines are single traces. In the bottom panels, the black line shows the mean firing rate and the red line shows mean plus one standard deviation of firing rate.

2.4.2.2 Avgis Hadjipapas' group (University of Nicosia)

Background: The aforementioned phenomena such as the shift in oscillation frequency (observed in monkey LFP and human MEG) and the decay in power (observed in monkey LFP but not human MEG) can be accounted for by a number of different mechanisms in the underlying neuronal network, even when this is only modelled in a spatially undifferentiated fashion. We are currently focusing on devising models that can replicate the monkey data. Our models do replicate both the increased gamma frequency and the power decay with increased input. However, we have observed these phenomena in a network where the rhythm is produced by excitatory neurons that fire more frequently than inhibitory neurons. This is not necessarily a realistic scenario; based on existing neurophysiological literature one expects relatively sparse firing of excitatory neurons [38–40] and faster firing of inhibitory neurons[41], which also are the ones that typically generate the

gamma rhythm. Thus two very different mechanisms could account for the same phenomena in the aggregate signals.

Aim: The fact that different model architectures can lead to the same network output, show that additional constraints have to be built in into the model. To that aim, we are currently combining the a-priori principles of Pyramidal Interneuron Gamma networks (PING) with constraints from empirical data. We expect that this will help to identify the most likely underlying mechanism.

Results: The first stage of the Task involved building a realistic PING (Pyramidal Interneuron Gamma Model) model that could on the one hand replicate key experimental findings such as a power decay with increasing contrast (afferent input) and on the other hand work in a biologically- realistic regime for instance in terms of neuronal firing rates. The approach that we undertook was to guide the model by contemporaneous analysis of empirical macaque data. Although further validation steps are necessary (see UM task), preliminary analysis suggests the existence of a class with much lower spike rate (E) and another with a much higher spike rate (I). We constructed a model that like empirical data contained sparse-firing neurons, whereby model inhibitory neurons fired more frequently than model excitatory neurons and generated the rhythm. The model when perturbed by increasing afferent input, exhibited the core characteristics of the empirical data, that is,(1) a monotonic increase in LFP frequency, (2)a non-monotonic LFP power modulation with decay at high inputs, (3)a modulation of spectral asymmetry and (4) a largely non-saturating increase in average unit firing rate. When the fully validated classification of E and I classes becomes available from the work at UM, we will be in a position to fully update and empirically validate the model.

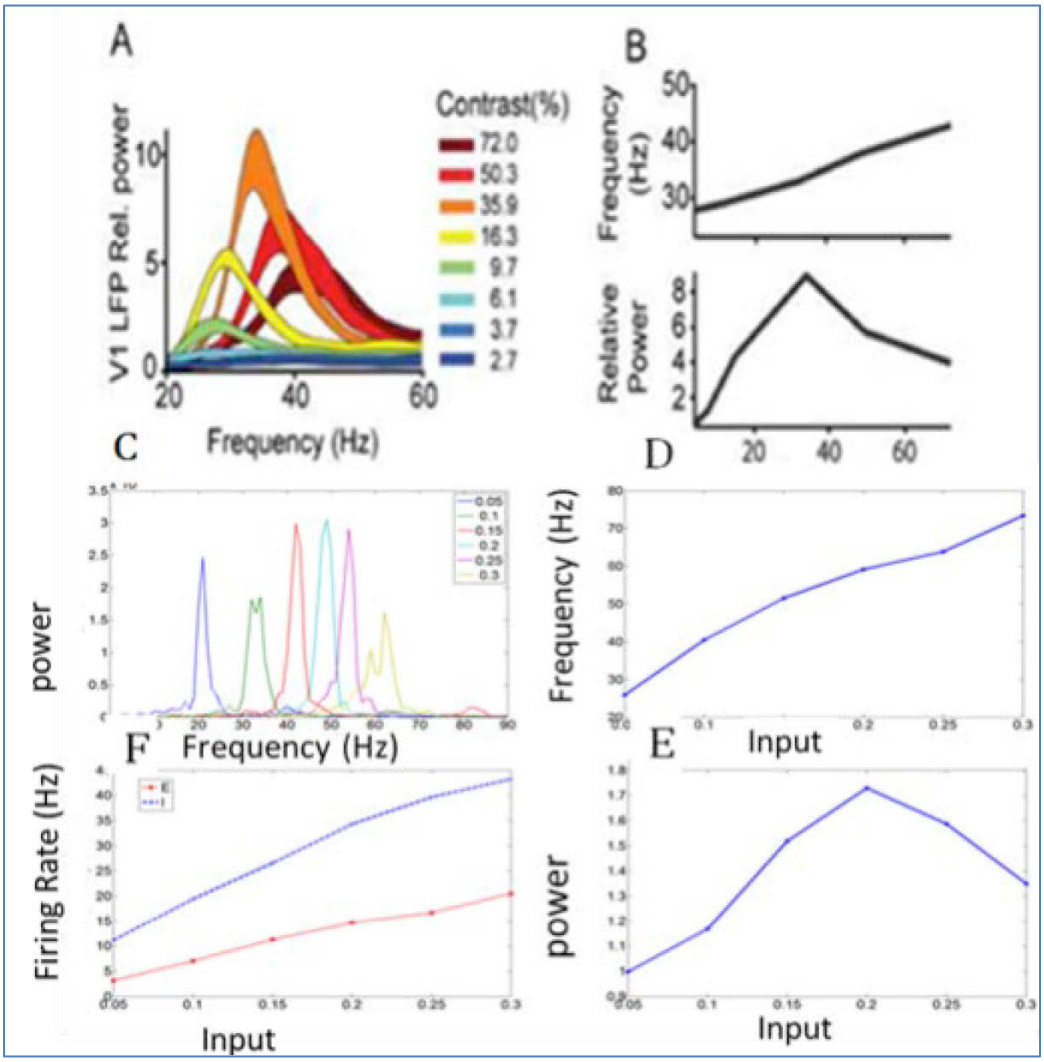


Figure 9: Panels A-B - Data from macaque visual cortex. Panels C-F - Output of PING model with 80 excitatory and 20 inhibitory neurons

Panel A. stimulus related power spectra as a function of grating contrast. Panel B. Top. Peak Frequency as a function of contrast. Note the monotonic shift in peak frequency with grating contrast. Bottom. Relative Power as a function of contrast. Note the non-monotonic change with contrast and decay at high contrasts. Panel. C. Power Spectra of the PING Model LFP as a function of excitatory input to the network. Panel D. Mean frequency of model LFP as a function of input. Note the similarity to LFP data. Panel E. LFP power as a function of input. Note non-monotonic relationship as in the real data. Panel F. Average firing rates for inhibitory (blue) and excitatory (red) neurons as a function of input. Note the non-saturating function of input. Also note the higher rates in I-neurons.

2.4.3 Plans for the next six months

Peter de Weerd's group (UM) will focus on finalising spike sorting and clustering into putative excitatory and inhibitory neurons. Then, we will extract the statistics of spike-LFP phase-coupling as a function of contrast. We plan to extract within and between-class population correlation measures as a function of contrast. Furthermore, we will quantify neuronal variability in putative excitatory and Inhibitory neurons. Finally, we plan to perform explorative data analysis in terms of layer- specificity of unit and LFP responses to varying contrast. All these steps are necessary to put us in a position to model columns, and then expand the model into a topographical model composed of interacting juxtaposed columns.

Avgis Hadjipapas' group (UNIC) plan to further constrain and validate its PING model with the results of the analysis of empirical data from the macaque visual cortex. We then plan to use the validated model to investigate the underlying mechanism of the gamma rhythm disruption and resulting power decline at high contrasts and its stabilisation for intermediate contrasts. One hypothesis we will test is that the power decline results from a primary (functional) decoupling among inhibitory neurons resulting in a desynchronisation among excitatory neurons. Finally, we intend to explore models that incorporate layer-specificity. This work contributes to the ultimate goal (after 30 months) of building a transfer model, because an empirically validated columnar and ultimately laterally expanded topographic model is expected to generate large scale interactions that may explain divergences between human MEG and monkey LFP/spiking responses, thus suggesting detailed functional similarities in cortical architecture in human and monkey V1.Alternatively, the observed differences in humans compared to monkeys may require model adaptations to account for them, suggesting significant differences in architecture among the two species.

2.5 Development and validation of brain network models constrained by realistic physiological phase lags and interaction time delays (T3.1.5)

2.5.1 Research goal

The objective of this Work Package is to understand the dynamic nature of spontaneous brain activity through both electrophysiological recordings and biologically realistic brain network models with concurrent oscillations in multiple frequency bands.

Utilising an up-to-100-subject cohort of intra-cranial stereo-electroencephalography (SEEG) data, we aim to rigorously characterise the spontaneous spatio-temporal patterns in resting- and task-state data and to achieve a comprehensive map of 1-500 Hz dynamics as well as inter-areal connectivity and phase/time lags among cortical areas. Comparative M/EEG and fMRI data from healthy controls will tie SEEG results with non-invasive brain imaging. Beyond the cohort size, the two key advantages of these SEEG data are sub-millimetre-accurate localisation of the electrode contacts in cerebral tissue and the consequent ability to use local white-matter contacts are "silent" references for nearby grey matter contacts. This, for the first time, will yield gold-standard-quality measurements of not only inter-areal interactions but also of the associated phase and time lags that cannot be measured accurately either non-invasively or even with bipolar SEEG.

The knowledge of neuronal communication delays is fundamental for any biologically realistic modelling of brain network activity. We will elucidate precisely the structural and anatomical constraints (connectivity, time delays) imposed upon a network, which lead to the emergence of self-organised brain pattern dynamics with the ultimate goal of understanding how distributed information is integrated for the constitution of human cognition. Utilising then lag estimates from SEEG together with MRI-based structural connectivity maps for biologically realistic large-scale brain network simulations, we will account for how coherent oscillations emerge and how they integrate robustly distributed information. We will then validate the theoretical findings against SEEG and magneto-/electroencephalography (M/EEG) data obtained from resting-state and two visual task conditions.

2.5.2 Main achievements

2.5.2.1 Satu and Matias Palva's groups

The Satu- and Matias-Palva groups (UH) have reached all internal goals planned for the first six months of the project (April 2014-Sept 2014, with funding from June 2014).

Experimental design and stimulus generation: We have defined the experimental paradigm to comprise a resting-state session and two continuous performance tasks, and implemented the task protocols. The tasks are a visual threshold-stimulus detection task (TSDT) and a visual threshold-discrimination task (TDT), where long-range temporal correlations (see Figure 10A) are governed by those in attentional[42] and putatively perceptual[43], respectively, systems, while both tasks yield event-related dynamics related to detection and discrimination at the behavioural level. To validate the tasks and the pre-task calibration procedure, we are acquiring psychophysical test-retest reliability data ($N_{planned}$ = 18, $N_{current}$ = 14) of which the interim analysis shows that test-retest correlations among behavioural scaling exponents (Figure 10B) were significant (r_{TSDT} = 0.55, $p < 0.05$; rTDT = 0.77, $p < 0.001$, Figure 10C). All participants also undergo a neuropsychological screen. Software for generating these experiments in MEG and fMRI will be released and they could be valuable in the Localiser Toolkit.

Pilot data and preliminary analyses: The manuscript describing the SEEG electrode localisation algorithm is provisionally accepted to BMC Bioinformatics. Likewise, the manuscript describing the novel approach for closest-white-matter-contact referenced SEEG recordings that will be used in HBP is provisionally accepted for publication in NeuroImage (Figure 10D). In this paper, we also describe the first comprehensive analyses of phase and amplitude interactions among human cortical sources and illustrate that our SEEG contact localisation yields distinct signals from deep and superficial cortical layers. Initial connectomics analyses of resting-state data (N_{SEEG} = 22, N_{MEG} = 14 after rejections) have also been completed. These analyses showed that neuronal-avalanche like activity cascades preferentially propagate along specific cortical pathways (Figure 10E) are characterised by power-law size distributions in both MEG and SEEG (Figure 10F). This avalanche propagation matrix (APM) at critical scaling (see Figure 10E) is significantly and frequency-dependently correlated with the electrophysiological connectome of phase-phase (PP) and amplitude-amplitude (AA) coupling (Figure 10G). These observations lay the basis for both the identification of communication delays and the comparisons between model dynamics.

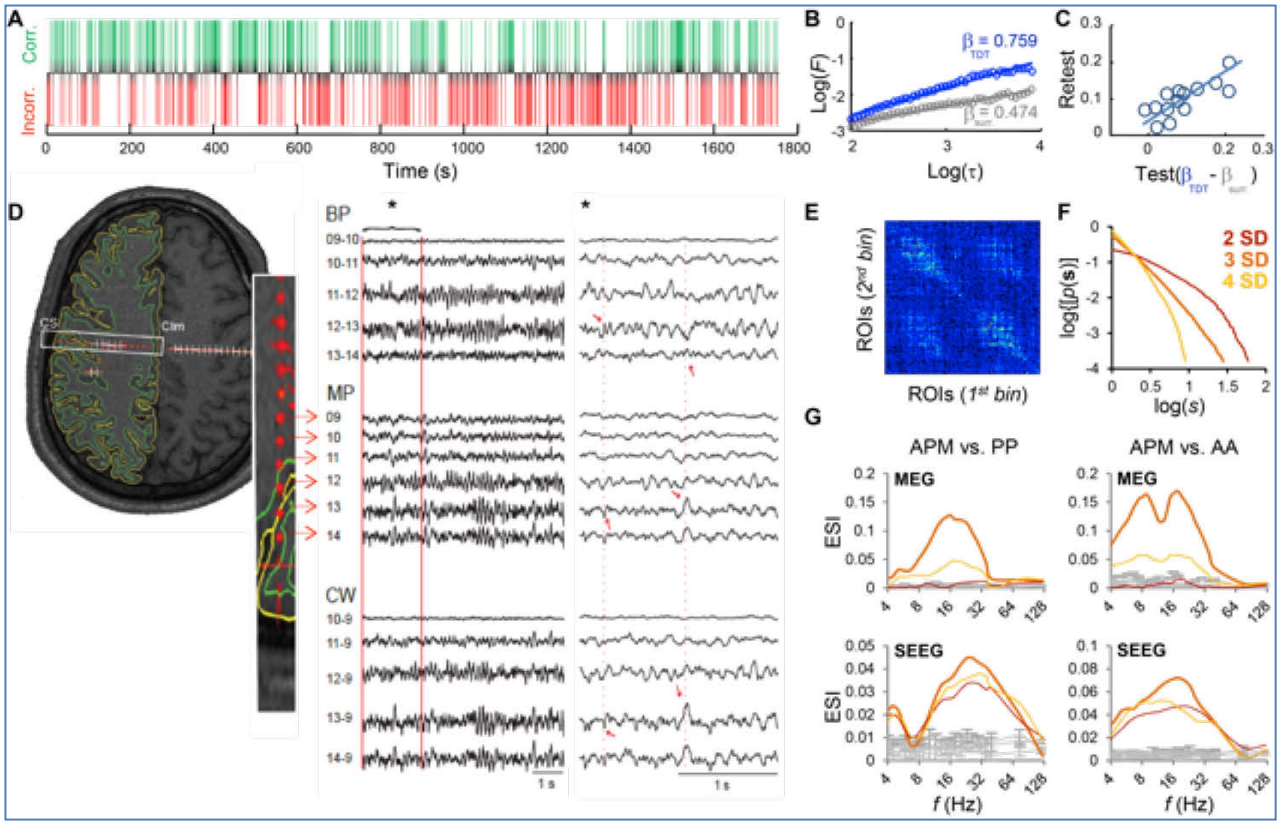


Figure 10: Experimental paradigm and electrophysiological methods for probing critical brain dynamics and its behavioural implications

(A) Behavioural responses in a Threshold Discrimination Task (green correct, red incorrect responses from a representative recording) exhibit emergent (B) long-range temporal correlations (blue, single subject; grey surrogate data) in a detrended fluctuation analysis (C) of which the scaling exponents Retest Test reliable at the group level (N = 14, r = 0.77, p < 0.001). Similar trait-like dynamics characterise Threshold Stimulus Detection Tasks1. (D) In stereo-electroencephalography (SEEG) recordings from human patients in pre-surgical evaluation yield electrophysiological data from 152 ± 20 electrode contacts in 14 ± 1.9 shafts / subject. The neuroanatomical localisation of these contacts can be identified with sub-millimetre accuracy. This opens the possibility of referencing grey-matter contacts to those in closest white-matter (CW), which eliminates the massive volume conduction effects in monopolar (MP) recordings but without the time and phase distortion inherent to bipolar (BP) recordings. (E) Neuronal activity cascades, 'neuronal avalanches', propagate along preferred pathways in MEG and SEEG (not shown) and (F) exhibit power-law size-distribution scaling in avalanche-detection-threshold-dependent manner. (G) At power-law scaling (orange lines), avalanche propagation (APM) in human cortex is similar (ESI) to the electrophysiological connectomes of phase-phase (PP) and amplitude-amplitude (AA) coupling in both MEG and SEEG.

2.5.3 Viktor Jirsa's group (SP3_SKPI-21)

Jirsa's group (AMU) is advancing with the model development as planned and has met the Milestones.

Context: The anatomical connectivity and the associated signal transmission delays define the *Space-Time Structure (STS) of couplings*. Any coherent oscillations within frequency-specific large-scale networks will necessarily have to respect the constraints imposed by the STS of the couplings; thus an understanding of the influences of time delays upon the synchronisation behaviours is essential. Most theorems on the influences of time delays upon network dynamics make either the assumption of one single common delay, temporally (but not spatially) distributed delays, or allow for maximally two delays. There is no theorem on how multiple different spatially distributed delays may shape an emergent global network dynamics. As a consequence, in absence of a better approach,

these delays are mostly ignored, though the community clearly acknowledges its existence and importance.

Approach: Initial studies of the STS found in DTI data of ten human subjects show a bimodal distribution of tract lengths (see Figure 11). We hypothesise that such multimodal organisation in the STS imposes limits upon the synchronisation behaviour of an oscillatory network. To illustrate this point, we develop a network based on simple coupled oscillators of Kuramoto type and implement couplings c_{ij} with a bimodal STS, time delays τ_{ij} and frequencies w_i. Then the dynamics reads

$$\dot{\varphi}_i(t) = w_i + \sum_j c_{ij} \sin(\varphi_i(t) - \varphi_j(t - \tau_{ij}))$$

We run first simulations and test the networks convergent/divergent synchronisation properties when the STS is altered.

First results: In Figure 11, we demonstrate that there exists a bimodal distribution in the space-time structure (anatomical connections and time delays from DTI data) of the couplings in the human brain. We further show selective multi-cluster synchronisation and identify the corresponding subnetworks, which are synchronisable, as well as the conditions upon the network parameters, under which synchronisation is achieved. These results provide first proof of concept that it is indeed possible to uncover the dynamic structures hidden in multi-delay network systems.

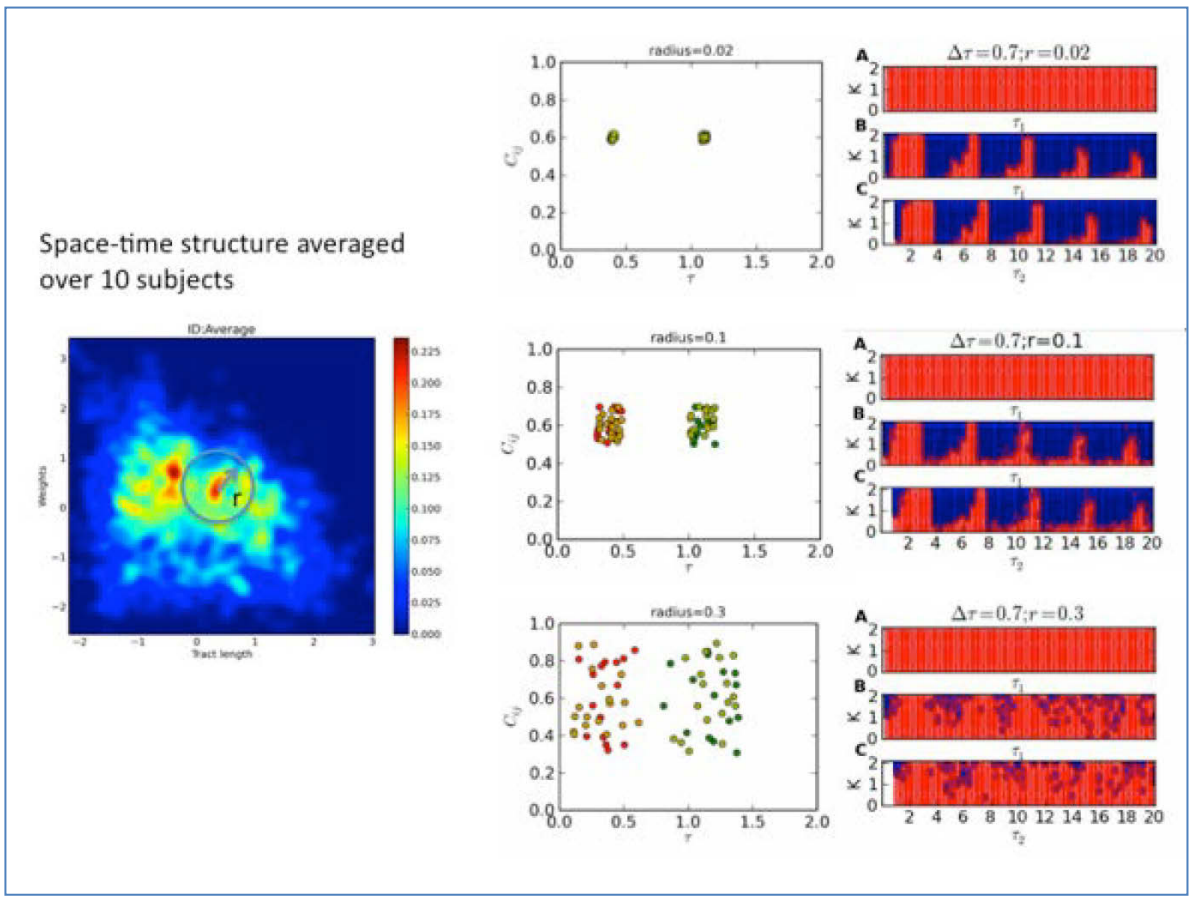


Figure 11: Synchronisation due to space-time structure of couplings

Left: Empirical space-time structure of couplings demonstrates bimodal distribution. The parameter r measures the dispersion of one peak. Middle: Simulated space-time structures of couplings with varying dispersion r. Right: Synchronisation plots of the full network (top), subnetwork with short time delay (middle) and

subnetwork with long time delay (bottom). Red indicates asynchrony, blue synchrony. Each triple corresponds to the adjacent space-time structure for varying r.

2.5.4 Plans for the next six months

Following the proof-of-concept analyses carried out so far, the *development stage* is further advanced by the mapping of phase lags and interaction time delays in the SEEG functional connectome. We will also address the relationship of axonal pathway lengths, fibre densities, and the space-time structure therein with the time and phase lag distributions estimated from SEEG data. Neuronal interaction delays will then be used as realistic constraints for the full brain network model that will be developed and implemented on the simulation platforms. Simulations will be performed to systematically map out network behaviours as a function of connectivity and delays. We will also address whether synchronisation regimes and dynamics identified with the model (see Figure 11) can be observed in the corresponding SEEG data.

3. Motivation, Decision and Reward (WP3.2)

3.1 Mapping and understanding the neuronal circuits involved in decision making, confidence and error correction (T3.2.1)

3.1.1 Research goals

The purpose of this Task led by Mariano Sigman is to propose a unified framework for the study of confidence in a computation or a decision.

Understand how human subjects and, as it turns out, several animal species can assign a rather accurate degree of confidence to their judgments is essential for any model of the human brain. Accurate confidence estimation fits with the idea that the brain operates as a statistical or "Bayesian" device that can compute with statistical distributions. If validated, this idea would put very strong constraints on any model of brain architecture. Yet confidence judgment is also subject to strong biases and inaccuracies. Characterising the limits of confidence computations is essential in order to understand what algorithm the brain uses to manipulate statistical data.

In the present work, the computational approach is used to quantify, in a solid mathematical framework, the estimation of confidence made by human subjects or non-human animals in a variety of cognitive domains. This framework builds on concepts from statistics and decision theory, in particular Bayesian decision theory. Computationally, it allows the distinction between different attributes of confidence (biases, accuracy, criterion…); experimentally, it detaches the notion of confidence from the single domain of explicit reports by subjects: implicit measures of confidence (choices, reaction times, etc.) can be investigated in human infants and non-human animals. The framework should thus be tailored 1) to investigate a range of behavioural and neural signatures of confidence, across different tasks, species and levels of development, 2) to identify from which level of computation and information processing they emerge and 3) to characterise the difference between these signatures.

3.1.2 Main achievements (SP3_SKPI-04)

The initial phase of the work was supported by an in-depth review the knowledge in the field. Based on this review, the work was divided along three lines:

1) Together with Mariano Sigman (CEA/Buenos Aires) and Florent Meyniel (CEA), we identified a paradigm to serve as a localiser for confidence in the brain, using functional MRI. This paradigm is intended to factor out the determinants of choice and confidence in perpetual discrimination task. Perceptual decisions are often used to investigate confidence because they allow a strict control of the level of evidence provided to the subjects. In addition, the paradigm selected as a localiser was developed along analytic tools which allow the identification of partially uncorrelated components of choice an confidence, thereby potentially isolating what is specific to confidence[44] as opposed to choice . This localiser is presented in the Annex D.

2) Florent Meyniel (CEA) developed a new paradigm to investigate aspects of confidence that are currently under-explored. This work comprises a theoretical aspect: an optimal, information-theoretic model was implemented to guide the analysis of subjects' data; and an experimental aspect, with the completion of a behavioural study to validate the approach. Two fMRI pilots were acquired and are currently analysed. This protocol is described in more detail below.

3) Together, Florent Meyniel (CEA), Mariano Sigman (CEA/Buenos Aires) and Zach Mainen (FCHAMP) started to draft a review about confidence in the brain. The aims of this

review are to provide quantitative and unified foundations for the study of confidence; to review the evidence that some forms of confidence are processed by humans and non-human animals, across a variety of tasks, behavioural and neural measures; to characterise these forms of confidence and to investigate how they might be implemented in the brain. This draft is presented in the Annex C.

3.1.2.1 Experimental work

3.1.2.1.1 Rationale of the study

Confidence goes hand by hand with the converse notion of uncertainty. There are several origins for uncertainty. First, our means to perceive the world are noisy and ambiguous. Second, our world is intrinsically uncertain: processes are usually stochastic rather than deterministic. And our world in not only stochastic, it is also changing randomly, defining nested levels of uncertainties. Subjects may try to infer these stochastic characteristics to improve their decision. There is seldom a single good estimate of such characteristics, instead, a range of estimates with varying degrees of likelihood. In other words, we can be more or less confident in what we learn from the world.

So far, confidence has been investigated following two major lines: 1) perceptual confidence, e.g. by presenting subjects with noisy percepts[45–48] 2) self-monitoring of performance, using error-detection tasks, in which the error rate is experimentally increased by using a time pressure or degraded stimuli[49] and performance rating tasks, in which subjects rate their levels of confidence, e.g. when answering a general knowledge question[50–52]. The issue of whether and how confident we are about what we learn is comparatively unexplored. Typical learning models consists of rules determining how the observed data should be used to update an internal estimate[53–56], but crucially, this estimate is a single value, not a range of values, which excludes the notion of confidence in what is learn.

3.1.2.1.2 Paradigm

To bridge the gap between the research on confidence and learning, we adopted a task in which subjects guess the hidden characteristics of a particular process, described in the figure below.

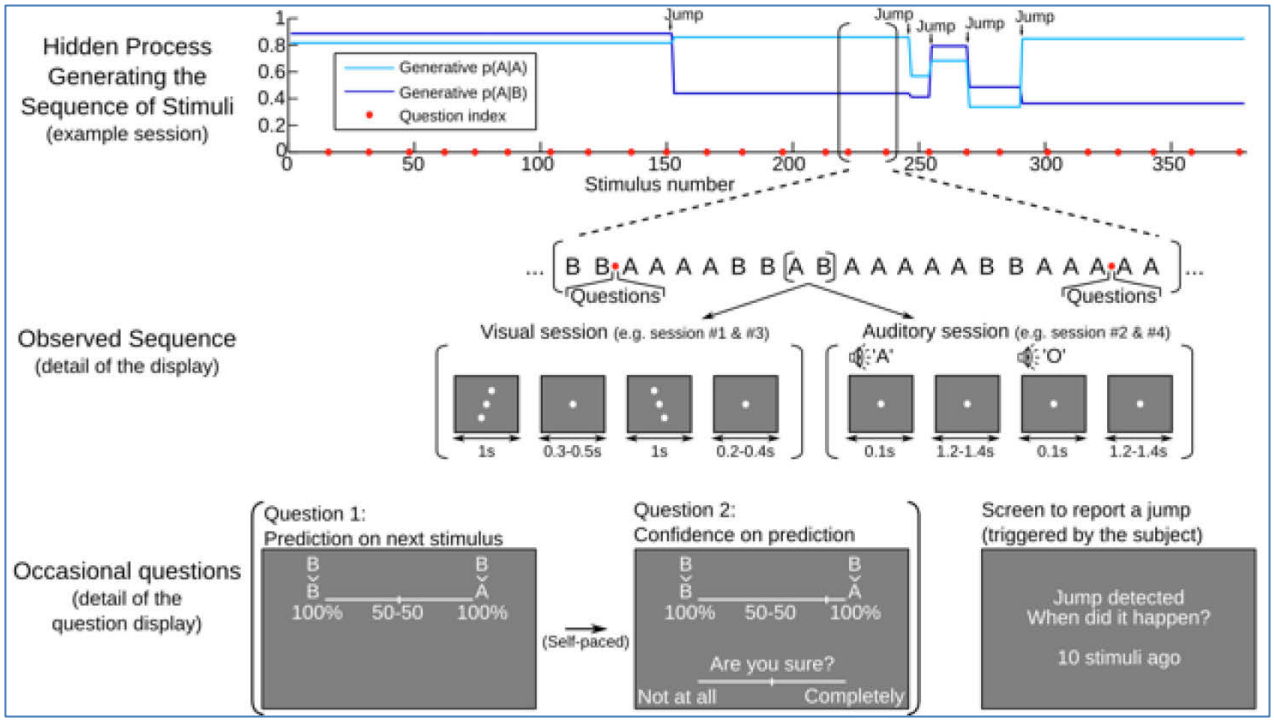


Figure 12: Task design for T3.2.1 - Mapping and understanding the neuronal circuits involved in decision-making, confidence and error correction

Subjects were exposed to sequences of binary stimuli (say, A and B) generated as follows. The sequence depended on the transition probabilities between stimuli A and B; these transition probabilities themselves were stable piece-wise: they were constant only for a limited time and changed abruptly and randomly, delineating 'chunks' in the sequence separated by 'jumps'. Subjects had (1) to report the occurrence of jumps, and occasionally they were jointly asked (2) to estimate which stimulus should come next (i.e. guess the transition probabilities in sensory stimuli) (3) to report their degree of confidence in those estimates.

Distinct modalities (visual vs. auditory) are used in separated session to disentangle in the fMRI study, which confidence-related processes are modality-specific or more abstract. Jumps were introduced in the sequence to prevent subjects from converging to a single estimate, but instead to foster the estimation of a range of potential and competing values, sufficiently sustained in time to reveal the distributional nature of subject's inference.

To guide the analysis of the subjects' data, we developed an Ideal Observer model[57]. Knowing that the sequences are generated by transition probabilities between stimuli and 'jumps' of these transition probabilities, this Ideal Observer provides the best inference of jump occurrence and transition probabilities. Thus, it provides predictions and confidence levels that can be compared to subjects' estimates.

3.1.2.1.3 Research results

In the past 12 month, Florent Meyniel (CEA) implemented the Ideal Observer model, the behavioural paradigm and acquired data from a group of 18 subjects. These data are now fully analysed, the results are about to be submitted for publication, together with Stanislas Dehaene (CEA). The behavioural study validated the paradigm and revealed that subjects not only made reasonable predictions based on what they learned, but they also maintained an accurate sense of confidence in these predictions. The dynamic of these predictions and confidence suggests that prediction and confidence are both derived from the same inference process, and this inference process shares several properties with the probabilistic inference.

These results were discussed at several occasions with Mariano Sigman (CEA/Buenos Aires), and in two meetings, one with Zach Mainen (FCHAMP), Rui Costa (FCHAMP) and Alejo Salles (Buenos Aires), at the Champalimaud Institute (Lisbone, April) and a joint SP3-SP4 meeting (Paris, June).

Based on the successful behavioural study and the powerful computational account provided by the Ideal Observer, Florent Meyniel (CEA) devised an fMRI version of the study and acquired data from two pilots.

3.1.3 Plans for the next six months

3.1.3.1 Localiser

- Florent Meyniel will implement an fMRI version of the paradigm previously developed by Mariano Sigman's team and acquire two pilots (November 2014)
- Florent Meyniel and Mariano Sigman will analyse the data from the two pilots

3.1.3.2 Experimental work

- Florent Meyniel will analyse the data from the 2 fMRI pilots (September-November 2014)
- Depending on the pilot results, a cohort of 20 subjects will be acquired (November 2014 – January 2015)
- Florent Meyniel will visit Mariano Sigman in Buenos Aires (December 2014) to work on the pilots from the localiser and the review

3.1.3.3 Review work

- Split the work among Florent Meyniel (CEA), Mariano Sigman (CEA/Buenos Aires) and Zach Mainen (FCHAMP) and start writing (September 2014 – February 2015)

3.2 Mapping and understanding the neuronal circuits involved in motivation, emotion and reward (T3.2.2) SP3_SKPI-05

This Task is currently not funded and will start at Month 13. It is led by Mathias Pessiglione (ICM).

The general objective is to build a neuro-computational model of the mechanisms that motivate the behaviour. These mechanisms include:

- Assigning subjective values to potential world states
- Comparing values so as to select the best option
- Aligning the direction and intensity of behaviour to the selected goal
- Updating option values based on the experience of behavioural outcomes

To probe the neural correlates of these basic mechanisms, a set of short behavioural tests will be developed. The idea is not to imagine new tests but to optimise existing tests in order to target key dimensions of motivation. These tests will be used in:

- Neuroimaging studies – We will employ both fMRI in healthy subjects and intra-cranial recordings in epileptic patients to characterise the spatio-temporal pattern of activity underlying motivational processes.

- Clinical studies – Patients with neurological or psychiatric conditions will perform the same tests to establish causal links between neural perturbation due to disease or treatment and behavioural deficits.

Results will be communicated in the form of review papers. These papers should:

- Position a particular test with regard to the relevant literature
- Suggest a computational account of the targeted mechanisms
- Provide links from computational variables and processes to neural entities

In the end, the neuro-computational models developed in this Task should enable predicting behavioural disorders from simulated neural perturbations, and reciprocally, inferring neural dysfunction from observed behavioural deficits.

3.3 Dissecting the brainstem modulation of cortical decision computations (T3.2.3)

3.3.1 Research goal

We intend to determine how phasic modulatory effects from the brainstem locus coeruleus (LC) noradrenaline (NA) system shape decision processes in the cerebral cortex. To this end, we will combine recordings of pupil diameter (a proxy of central neuromodulation), multimodal neuroimaging (fMRI and MEG), and algorithmic modelling of choice behaviour. Our results will constrain biophysically detailed circuit models of the decision-making.

3.3.2 Main achievements (SP3_SKPI-20)

Tobias Donner's group (UvA) has collected the first fMRI data sets using three elementary decision-making tasks known to involve phasic LC-modulation: yes-no visual contrast detection, an auditory oddball task, and a bistable visual perception. We used high resolution (in-plane resolution: 1.78 x 1.78 mm2), while covering the LC as well as key nodes of the cortical network processing these decisions (visual/auditory, parietal, and prefrontal cortex). We tracked respiration and heart rate, and removed the associated physiological artefacts by means of RETROICOR. We have sorted trials in high vs. low pupil trials based on single-trial pupil dilation values. We delineated the LC in individual subjects by means of neuromelanin-sensitive structural MRI and retinotopic cortical areas (V1-V4, IPS0-2, etc.) by means of retinotopic mapping. In parallel, we have begun to implement the effects of neuromodulation on cortical decision computations by extension of the leaky competing accumulator (LCA) model of decision-making. The model accurately reproduces the relationship between neuromodulation and choice behaviour (in yes-no and 2AFC tasks) that we observe empirically with pupillometry.

Initial problems with individual LC-localisation have been solved.

Andreas Karl Engel's group (UKE) has collected the first complete whole-brain MEG data set using a motion discrimination task with simultaneous pupillometry. We have parameterised the psychophysical results in detail (including quantification of learning-induced improvements in performance) and calculated decision-related pupil dilation responses for all subjects. We have also established diffusion model fits to reaction time distributions from the motion discrimination task.

3.3.3 Plans for the next six months

Tobias Donner's group will estimate LC-specific fMRI responses (extracted from individually defined LC-ROIs) and compare a number of established fMRI signatures of decision-related

activity in visual and parietal cortex between high and low levels of LC activity and decision-related pupil dilation. In the subsequent phase, we will collect pilot data at 7T with the hope to move even higher resolution (goal: 1x1 mm2 in-plane resolution). We will collect data from the classical 2AFC motion discrimination task and begin to fit different algorithmic models (drift diffusion model, LCA, LCA with neuromodulation, etc.) to the behaviour in the yes-no and 2AFC tasks. The first data sets and first draft cognitive architectures are planned for month 21 (MS276).

Andreas Karl Engel's group will collect the corresponding data set for a yes-no contrast detection task (including confidence ratings) and a spatial 2AFC contrast detection task. We will then assess the following established MEG signatures of decision-related cortical activity in all data sets: Gamma-band activity in visual cortex, sustained fields in parietal cortex, and beta-band lateralisation in motor cortex. Contrasting the time course of these signatures between trials with high and low decision-related pupil dilation will determine whether neuromodulation affects cortical dynamics during, at the end of, or after decision formation. Further, we will assess the effect of neuromodulation on decision-related long-range interactions, in particular fronto-parietal beta-band coherence.

3.4 Characterise multi-scale brain architecture of decision related motivational states and values (T3.2.4)

3.4.1 Research goal

Human behaviour is guided by motivation, often defined as goal directed behaviour. To understand human behaviour we must understand how environmental cues (both rewarding and punishing) affect us and lead to specific decisions. The goal of this Task is to study these processes in humans during interactive motivational related scenarios. Two engaging tasks (PRIME and RCD) are performed in order to isolate stimuli perception, conflict resolution, anticipation and response to reward or punishment. The data is obtained using intracranial recordings (in subjects with epilepsy) so to obtain a rich set of multi-level recording. At the micro level, single cell firing patterns and Local Field Potentials (LFPs) are recorded. At the regional and inter-regional levels LFPs, intracranial recordings (ECoG) and fMRI will be used. This multi-level perspective will allow a novel and more complete description of motivational decision making. This description will assist in formulating a neurocomputational model of motivational decision making.

3.4.2 Main achievements (SP3_SKPI-18)

Talma Hendler's group (TASMC) joined the HBP in April 2014 after being granted within the competitive call. During this half-year, we prepared the protocols for data acquisition from epileptic patients. We adjusted our psychological paradigms to fit the multiple level recordings. Finally, we have begun to record data from patients implanted with depth electrodes. The data from one grid patient and two depth electrode patients has been collected and initial data analysis has been performed. These initial results were presented at the HBP educational meeting in Tel Aviv, June 2014 and at the HBP annual summit at Heidelberg. Progress is according to plan.

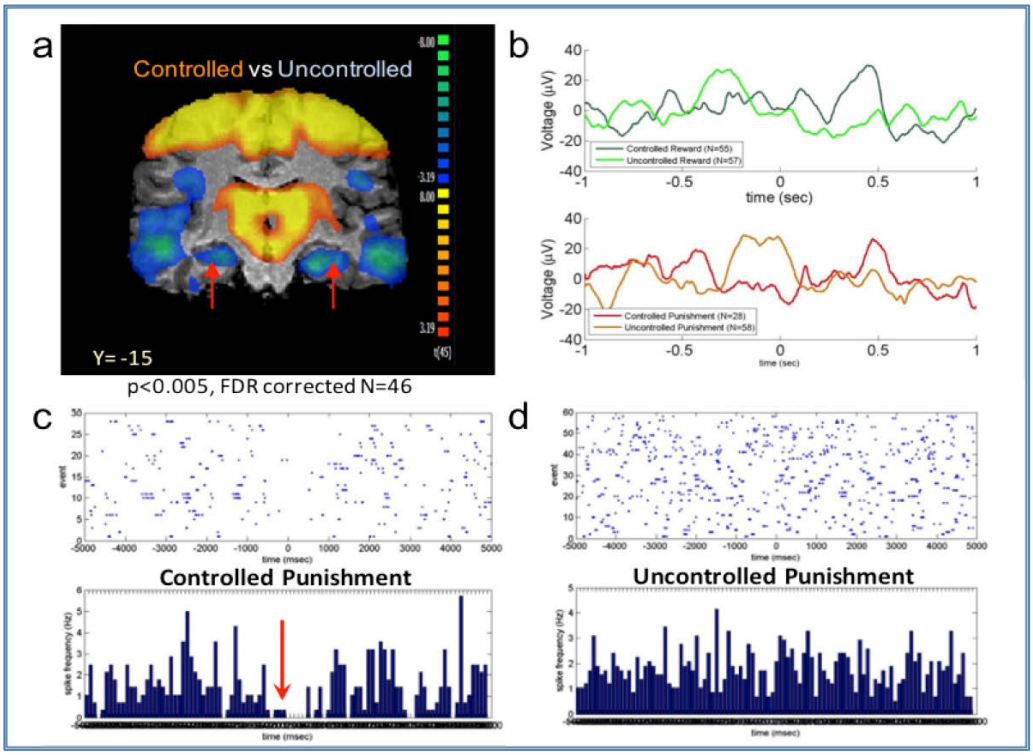


Figure 13: Example of multi-modal analysis in the right hippocampus of an epileptic patient using the PRIME paradigm.

(a) fMRI group analysis found higher activation for the uncontrolled vs. controlled conditions. (b) Event Related Potentials (ERP) analysis from LFP recordings from the right hippocampus. A positive peak at the time and before acquiring a punishment (bomb hit player) was found only for uncontrolled trials (bottom panel, orange) and not for the controlled trials (red). A reverse effect (with smaller amplitude) was found around 500ms after the punishment. Relatively similar effects were found for the reward conditions (top panel, green). (c-d) Raster plots of single cell firing for the control and uncontrolled punishment conditions. A decrease in firing rate was found only for the control punishment condition. Taken together, these results hint to decreased activation of the hippocampus in conditions where the stimulus contains an incentive value (controlled) vs. the condition in which only a hedonic value is apparent (uncontrolled). Future analysis will examine the dynamics of this phenomenon for modelling.

3.4.3 Plans for the next six months

In the next six months TASMC plans to continue the acquisition of multi-level data, acquiring data from three more subjects with intracranial recordings. Moreover, we plan to analyse this dataset and provide a report of region specific findings for the different motivational values (hedonic and incentive) and states (reward, punishment and goal conflict). Future work will expand this analysis to inter-regional findings.

4. Learning and Memory (WP3.3)

The aim of this WP is to map the neuronal circuits involved in human learning and memory systems. This WP is composed of 3 Tasks, T.3.3.1-3, described herewith, covering main aspects of human memory that are essential for future advances in human brain simulation and in neuromorphic computing and neurorobotics efforts.

4.1 Skills and Habits (T3.3.1)

4.1.1 Research goals

Skill acquisition and habit formation are critical to the ability to interact with the environment and to respond efficiently and appropriately to a rapidly changing world. Pathologies affecting these faculties are of great clinical and societal importance, e.g. in autism, schizophrenia. The same capabilities are of great importance for neuromorphic computing systems and neurorobotics. The goal of T3.3.1 is to characterise the neural systems and processes responsible for the acquisition of perceptual skills, motor skills, reward-reinforced and punishment-driven habits. The study will was intended to lead to high-level models of the neuronal circuits and of neuronally implemented algorithms responsible for the acquisition, persistence and modification of skills and habits to be developed in conjunction with SP4 in the second phase of HBP. However whether this longer-term essential goal will be implemented is hampered by the decision of the HBP governance to omit all studies for human systems neuroscience and cognitive neuroscience from the core of the HBP after the ramp-up.

4.1.2 Main achievements (SP3_SKPI-06, SP3_SKPI-07)

The group of Avi Karni (UHAIFA) studies cortical dynamics (repetition suppression, repetition enhancement, functional connectivity) as brain signatures of accumulating experience, plasticity and procedural memory consolidation in motor skill learning. Specifically, they test the conjecture that activity in area M1 is modulated by task repetition as a function of whether prior experience was afforded, and importantly, as signatures of overnight procedural memory consolidation. They have been re-analysing old and new data from an fMRI experiment addressing cortical dynamics in movement repetition for two movement sequences (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. They hence propose that M1 not only generates movements but also serves as a hub for a motor working memory (see T.3.3.3 below): wherein transient stabilisation of activity upon sequence repetition reflects short-term familiarity with a novel sequence of movements. A temporarily stabilised network in cortex and striatum may promote an integrated representation of the new movement sequence (i.e., the movement syntax). They also find that 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. A paper is currently in review.

In an additional project, motor cortex plasticity driven by visual input (action observation) is in progress. The behavioural data already suggest that executing and observing movements improve task performance and trigger skill consolidation processes. However, consolidation could be blocked by ensuing action but not by observation, indicating that skills acquired in doing or observing do not overlap in the brain. A paper is in preparation.

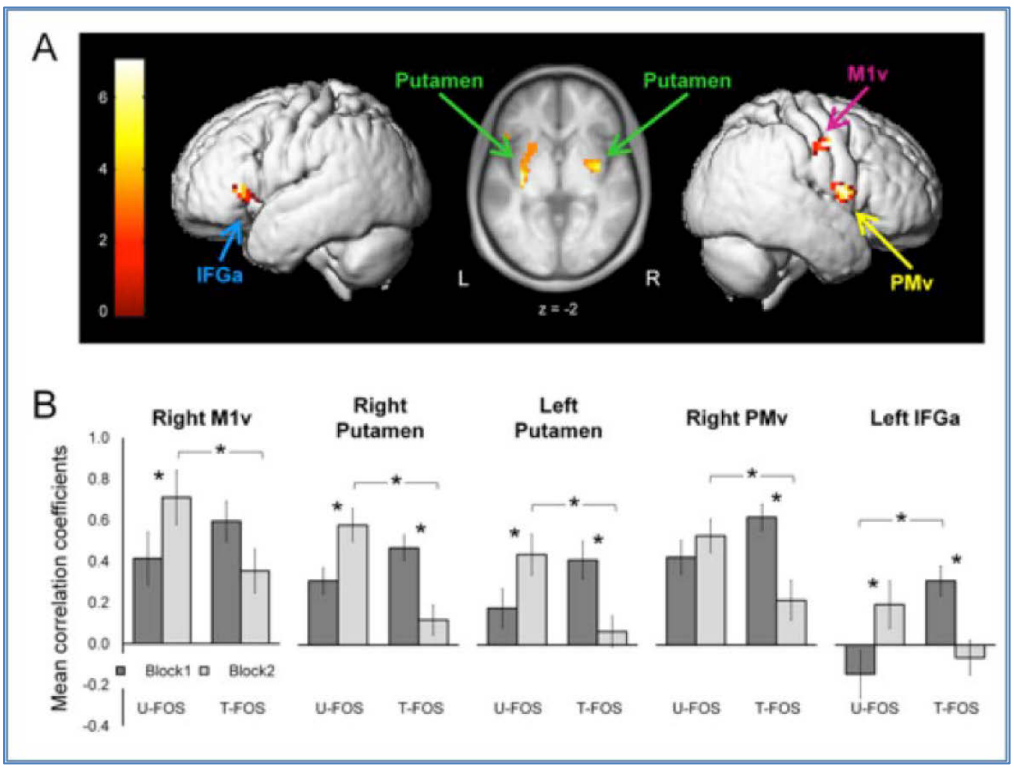


Figure 14: Illustration of work in T3.3.1 functional connectivity analyses using M1 hand area as a seed

(A) Areas wherein functional connectivity patterns, were differentially modulated by the repeated performance of the two sequences using M1 as a seed (interaction: [U-FOS: Block1 < Block2] X [T-FOS: Block1 > Block2]). Connectivity map of group effects is overlaid on the mean structural image of all subjects and the surface rendered from the subjects' mean structural segmented images. The map was threshold at $p < 0.001$. The colour-bar represents t-values. M1v – primary motor cortex ventral to the hand area, PMv – ventral premotor cortex, IFGa – anterior part of the inferior frontal gyrus. (B) Mean correlation coefficients between the M1 seed and each of the clusters, wherein connectivity with the M1 seed showed significant task by block interaction. Columns – mean Fisher-transformed correlation coefficients; bars – standard errors of means (s.e.m.). * – significant differences at 0.05 level corrected for multiple tests (i.e., a number of clusters) using Bonferroni adjustments.

4.1.3 Plans for the next six months

Setting up for the functional spectroscopy study; completion of imaging data collection in the aforementioned Experiment 2 and data analysis; finalising papers.

4.2 Memory for Facts and Events (T3.3.2)

4.2.1 Research goals

Memories for facts (semantic memory) and events (episodic memory) are key components of human memory. Furthermore, episodic memory is essential for imagination - mental time travel to the future requires the ability to simulate and compare alternative scenarios. This Task investigates brain mechanisms underlying encoding, consolidation, storage and retrieval of episodic memories, by combining custom-designed behavioural paradigms and human functional imaging (fMRI, MEG and EEG). The data, including cognitive rules and emergent boundary conditions on operation, stability and veracity of brain derived or brain inspired declarative memory systems, will contribute to the development of human brain models at various levels of resolution.

4.2.2 Main achievements (SP3_SKPI-08)

The research group of Yadin Dudai (WIS) has pursued its work on the boundary conditions of initiation of human episodic memory consolidation in the hippocampus during the first seconds after encoding[58,59]. This research has developed the first paradigm that permits teasing apart two potential computational modes of hippocampal function in human episodic experience[59]. One is of familiarity of target item attributes (probably reflecting pattern separation), the other encoding of novel item attributes (probably reflecting pattern completion in the hippocampus). These two modes are intermingled in real-life and their separation is essential for understanding and simulating brain circuits of episodic memory formation, actualisation and stability. A protocol for an fMRI localiser that permits identification of the engagement of distinct hippocampal sub-regions in these basic mnemonic functions is now further developed to fit is to the single subject level. The results of these studies performed in the context of the HBP have been so far published in three research papers[58–60].

The research group of Rony Paz (WIS) continued to focus on the effect of valence (i.e. positive vs. negative stimuli) on memory. Specifically, this group has performed experiments (combining methods of psychophysics, functional imaging and electrophysiology) that aim at understanding how memory generalises differently under positive vs. negative contexts. The group has identified the neural circuit that mediates the similarities and differences in the process. They further used this to develop a model for post-trauma and anxiety, and test it now in patients as well. The first part of the results is currently under review at a leading journal.

The research group of Jan Born (EKUT) aimed at modelling oscillatory EEG phenomena that underlie the passage of reactivated hippocampal memory information towards neocortical networks during slow wave sleep. Specifically, this passage is assumed to be supported by the hierarchical nesting of hippocampal ripples into thalamic spindles and neocortical slow oscillations (SOs). To this end, in a preparatory phase, we investigated the possibility to describe the sleeping brain within the neural mass framework, allowing for a cost efficient modelling of sleep EEG. With this approach, we could successfully explain the emergence of SOs as well as K-complexes during NREM sleep[61]. Next, we investigated the interaction between sleep spindles and SOs in the thalamo-cortical system (Schellenberger et al. in preparation), and are currently assembling a model of spindle-ripple event generation. This model shall be tested experimentally using data from ongoing simultaneous recordings of local field potentials and single unit recordings from hippocampus and neocortex in rats.

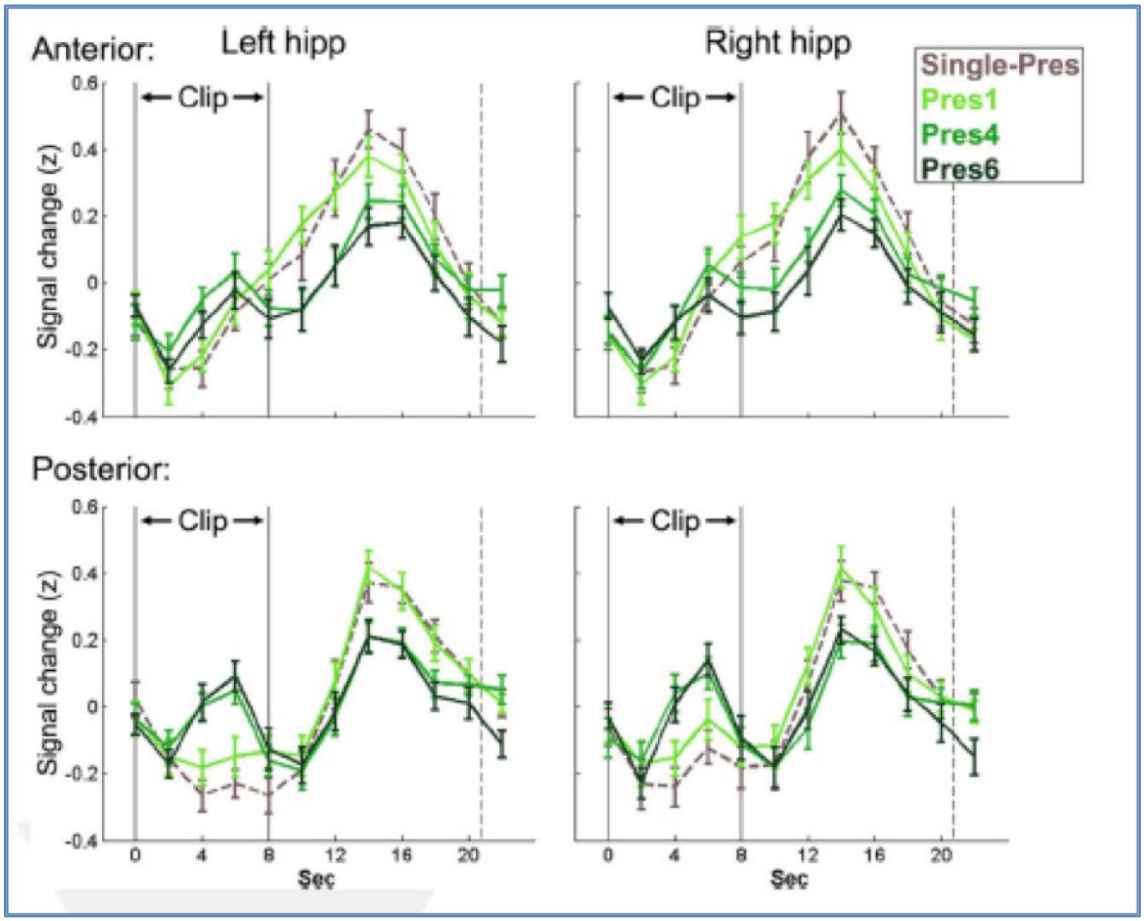


Figure 15: Illustration of work in T3.3.2, referring to teasing apart, for the first time, of two computational modes, encoding and retrieval, in the human hippocampus

In this figure, depicting BOLD signal change over time, repetition attenuates offline-locked hippocampal activity and induces an online response to a brief movie clip stimulus. Mean group BOLD signal (z-scored) during and following Single-Pres clips 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.

4.2.3 Plans for the next six months

Continue with further data analyses, and work with the neurocognitive constraints.

4.3 Working Memory (T3.3.3)

4.3.1 Research goals

Working memory (WM) maintains information over brief periods of time. This feature is required for goal-directed behaviour and allows us to act beyond the confines of the here and now. WM can thus be conceptualised as providing an interface between perception, long-term memory, and action. As such, WM is taxed by numerous laboratory- and everyday cognitive challenges. Previous human and primate research has shown that a network of brain regions implements WM, but the specific contribution of each network component remains unclear. In particular, it is unknown how the characteristic limited capacity of WM results from the functioning of different part(s)/network interactions. Also, a long-standing assumption of the relation between WM and consciousness has

recently been questioned and should be further investigated. T3.3.3 will contribute to the theoretical model of working memory developed by T4.3.2 ("models of working memory and the effects of attention").

4.3.2 Main achievements (SP3_SKPI-09)

A first draft of the cognitive architecture and localiser protocol for working memory was submitted at six months. While the localiser protocol remains unchanged, the group has shifted the focus of the locally performed fMRI experiment. Specifically, they have focused less on working-memory capacity limitation and instead explore the possibility of non-conscious working memory. For this, data collection and initial analyses has been completed and we are thus not behind schedule despite the change. The group has updated the KPIs accordingly.

4.3.3 Plans for the next six months

The Task team will continue with further data analyses, and work with the neurocognitive constraints.

5. Space, Time and Numbers (WP3.4)

Only the Space part ("identifying circuits for spatial navigation and spatial memory") is funded during the HBP ramp-up phase, and is therefore the object of the present report. The Time and Numbers parts were supposed to start only in the second phase (yet have not been retained in the new FPA).

5.1 Identifying and Analysing the Multi-modal Circuits for Spatial Navigation and Spatial Memory (T3.4.1)

5.1.1 Research goals

This work will contribute to the definition of the cognitive architecture of spatial navigation, one of the principal interests of SP3 (also with strong links to SP4). The modelling work will be potentially suitable for implementation on the large-scale network models developed in SP6 and in the neuromorphic computing devices implemented in SP9.

5.1.2 Main achievements

5.1.2.1 The cognitive architecture of spatial navigation

Neil Burgess' group (UCL) have made progress in the development of theoretical and computational models of rat behaviour in navigation tasks. We have focused on the implementation of a biologically plausible architecture that contains a sufficient number of areas strongly involved in navigation and spatial decision-making, but at the same time utilises a minimal number of parameters and simple but realistic learning rules. The idea being to capture the essential architecture required to support the behaviour of interest with the minimum number of free parameters, rather than under-constrained data-fitting with large numbers of parameters.

In particular, our model consists of four main areas: the hippocampus, the dorsal striatum, the ventral striatum and the prefrontal cortex.

The hippocampus has been linked to episodic memory: it is able to learn in one-shot sensory states (including rewards) and associate them to places. The population activity of place cells can then be used to infer the direction in which specific objects or places (e.g. rewarding sites) are. Unlike some existing models, we don't postulate that place cells form a connected graph through which it is possible to search for the optimal path, instead we utilise the hippocampus to build gradients that point towards the desired locations[62](SP3_SKPI-10).

The dorsal striatum is strongly involved in habitual learning. In our model it is part of an actor-critic architecture that gradually learns to associate sensory states to the actions (movements) that most efficiently take the agent to the rewarding site, via Q-learning[63]. In this case the ventral striatum is thought to be the critic[64] (SP3_SKPI-11).

The visual system of our simulated read is constituted of a one-dimensional colour/range sensor that provides distances along multiple lines of sight for different colours (ground, landmarks and boundaries have different colours). The visual input is then discretised into visual receptive fields and associated to different neurons. In this way it is possible to obtain a discretised activation of the sensory neural population and learn its connection to the ventral striatum (the critic) through TD learning.

5.1.2.2 Neuro-cognitive constraints for models of spatial navigation

In order to validate our model, we have replicated the experiments of Pearce and colleagues[65] where rats have to learn the location of a submerged platform in a Morris

water maze, whose location is kept constant relative to an intra-maze landmark while both platform and landmark are moved within the pool on every fourth trial. Here, in order to study the contribution of the two different navigation mechanisms (hippocampal and striatal), experimenters lesioned the hippocampus in one group of rats. Results (Figure 16) indicate that lesioned rats are still able to learn how to reach the platform, probably utilising the landmark as a reference point within the directional context of distal cues. Control animals on the other hand perform better when the landmark and platform remain in the same location, but have difficulties when the landmark is moved from trial to trial.

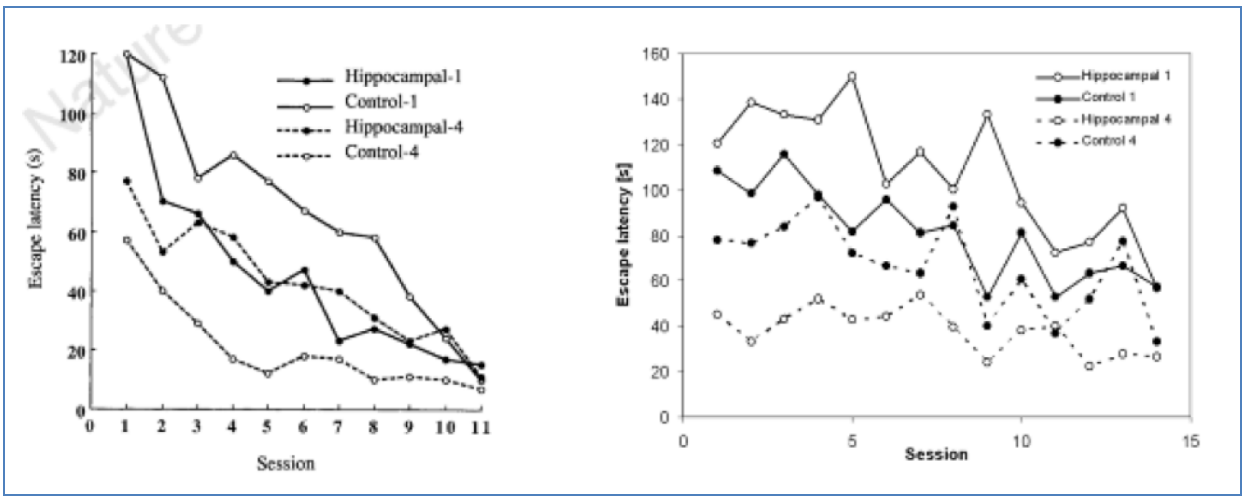


Figure 16: Results of navigation experiment in Morris water maze

In the right panel of the figure, we report the results of our and simulated agent for comparison.

5.1.3 Plans for the next six months

At Month 12, we defined the strategic data for modelling hippocampal navigation, based on existing literature, and have started simulating tasks regarding the hippocampal-striatal circuits for self-localisation and goal-directed navigation.

At Month 18, we will provide preliminary data on a small-scale, firing rate coded network model of hippocampal navigation: combining sensory inputs, path integration, and attractor dynamics, including the firing of head-direction, place and grid cells.

6. From Sensory Processing to Multimodal Perception (WP3.5)

The Tasks covered in WP3.5 are led by Yves Frégnac (CNRS, WP3.5 leader) and Brice Bathellier (CNRS) and implemented at the UNIC CNRS research centre. Models in the second phase of the project will be developed in collaboration with Andrew Davison (CNRS) involved in SP6.

The Task goals are to provide observations to constrain models of primary sensory cortical area aiming at reproducing the neural correlates of low-level perception in a unimodal (T3.5.1) or multimodal (T3.5.2) context. The integration strategy used here proceeds in several incremental steps. The first step is to extract, from existing data, generic principles that will define a first set of modelling constraints. This, in turn, will lead to the definition of strategic experimental paradigms. Their implementation will lead to novel data that will be used to test the validity of the coherence in knowledge integration obtained in the proposed models.

6.1 Neural correlates of unimodal perception and self-organisation of internal knowledge in mammalian primary cortical areas (T3.5.1)

6.1.1 Main achievements (SP3_SKPI-12)

Unified parametric protocols were proposed to search for the stimulus dependency of the temporal precision and sparseness of the neural code, for stimuli ranging from gratings to dense noise, including animated natural scenes with realistic eye-movements. The data obtained earlier with intracellular techniques have been fully processed and published. The same exact protocols have been repeated to study the evoked dynamics with mesoscopic measurements (local field potentials (LFP)s, multiple simultaneous recordings (MUA), and voltage sensitive dye (VSD) when possible). Data processing of the mesoscopic signals is still under way.

The intracellular data have been used to build a comprehensive data-driven model of V1 in the higher mammals (cat, ferret, monkey) applicable to man. Parametric search has been finalised to account for the receptive field dynamics of first-order Simple cells receiving direct input from the thalamus. Most functional properties of receptive fields as well as the context (stimulus statistics) dependency of the temporal precision of the neural sensory code are accounted for by the interplay of fast synaptic thalamo-cortical depression and generalised push-pull organisation of ON- and OFF afferents (feedforward and recurrent) during optimal (gratings) and non-optimal (natural statistics) regimes of sensory stimulation.

6.1.2 Plans for the next six months

The next six months will be dedicated to constrain a full-scale conductance-based model of V1 with the intracellular experimental data reported earlier. Mesoscopic data will be processed to be integrated at a later stage in the model. Experiments will be carried out to find neural correlates of the perceptual association field reported in psychophysics.

6.2 Neural correlates of unimodal and multi-modal perception in mammalian primary sensory areas (T3.5.2)

6.2.1 Main achievements (SP3_SKPI-13)

To provide data constraining the working architecture of sensory cortex, we have used two-photon calcium imaging to record population activity in visual cortex during auditory and visual stimulation using moving visual stimuli and intensity modulated sounds. After the successful acquisition of pilot data, we have observed that 300 neurons of mouse V1 could be recorded simultaneously with a signal to noise ratio sufficient to discriminate between played stimuli using population activity. The data shows that numerous neurons of mouse V1 respond not only to visual but also to auditory stimuli and that bimodal stimulation often leads to non-additive (non-linear) responses (Figure 17). We therefore have started to acquire a full dataset of V1 activity in the awake mouse during multisensory stimulation.

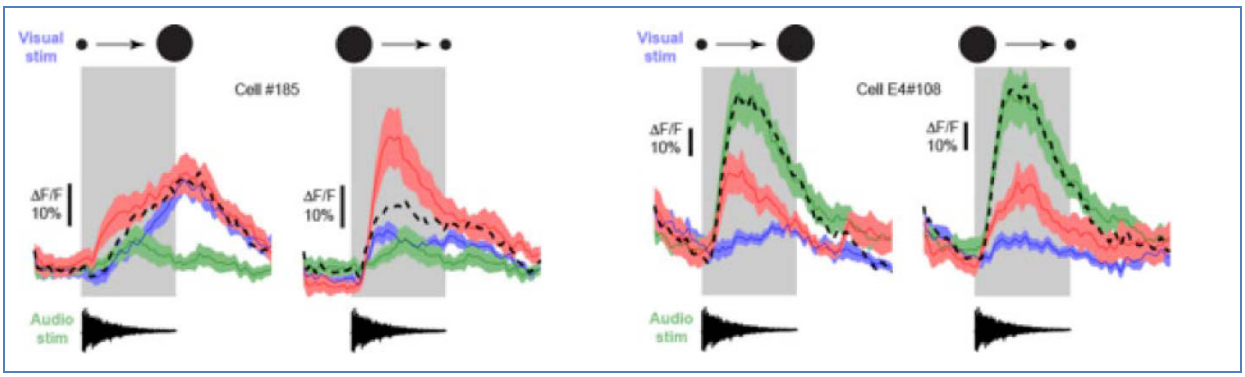


Figure 17: Calcium response of 2 example neurons to visual (blue), auditory and bimodal (red) stimuli

These neurons show responses to both modalities and non-additive bimodal responses.

6.2.2 Plans for the next six months

The next six months will be dedicated to acquiring a full set of visual cortex population activity data during auditory-visual stimulation in mice. In this way, we will achieve our Milestone M60 in month 18 as planned. We will also start detailed analysis of this data. In addition, we plan to perform visual cortex recordings during an auditory-visual discrimination task, which we have almost finished developing.

7. Capabilities Characteristics of the Human Brain (WP3.6)

7.1 Linguistic and Non-Linguistic Nested Structures (T3.6.2)

7.1.1 Research goals

The purpose of this Task led by Christophe Pallier (CEA) is to shed some light on the problem of how the human brain encodes tree structures – a neural code whose mechanisms are currently unknown, and which indeed has been hypothesised as perhaps unique to humans as opposed to other non-human primates. To clarify which areas are involved and how they operate, we plan 3 studies, differing in two ways. First, they present structures of varying degrees of complexity, starting from sequences structured by chunks and the transition probabilities between stimuli, to sequences in which the structure is nested. Second, these studies present linguistic and non-linguistic structures to isolate potential domain-specific brain processes. These studies are only partially funded by HBP.

7.1.2 Main achievements

We drafted the content of a review on the neuronal implementation of different aspects of sequence processing, ranging from the extraction of transition probabilities between stimuli, to nested structures. This draft is presented in Annex C.

We also achieved significant progress in our experimental work. Experiment 1 examines with fMRI how human extract the regularities of linguistic nested-structures, experiments 2 and 3 examine how humans extract the regularities of non-linguistic structure. Experiment 2 examines this capability with MEG data, using sequences structured by the transition probabilities between the stimuli. Experiment 3 investigates this capability with behavioural and fMRI data, using sequences structured by chunks and nested rules.

7.1.3 Experiment 1. Encoding of Syntactic Structures (SP3_SKPI-14)

Two fMRI studies searching for brain regions encoding syntactic trees are underway. In the first one, participants read very short sentences (2 to 4 words) that vary in the size of the tree postulated by linguistic analyses. These sentences were generated by applying a variety of syntactic operations (e.g. Wh, V or NP movement), allowing to ask whether some areas are specifically involved in some of these operations. We have completed the scanning of 20 human volunteers and are currently checking data quality and have started the data analyses. We have started a second fMRI experiment to investigate the processing of mathematical syntax, addressing the debate of whether there are domain-specific syntactic areas or a domain independent syntactic network. We manipulated the grammaticality and complexity of mathematical formula presented to mathematically-trained participants; and we also use a priming approach where some formula are presented in the context of others that share or not the same structure. The experiment has been designed, programmed, and 14 out of 20 participants have been scanned.

7.1.4 Experiment 2. Bayesian Modelling of Expectation Effects in Sequences (SP3_SKPI-15)

When exposed to any type of sequences, even random ones, humans search for regularities, like streaks and alternations. In line this idea, new observations may reinforce the current characteristic inferred from the sequence, or alternatively, be at odds with this estimate. When expectations are violated surprise signals are detected in the brain, e.g. as the amplitude of the P300 in EEG recordings[66], or with fMRI[67] and in the behavioural measures such as reaction times[68]. Existing data clearly identified that such expectancies

are impacted by local (corresponding to a short period of time) and global (corresponding to a longer period of time) characteristics of the sequence, such as the proportion of stimuli, and possibly, the transition probabilities between these stimuli[66,69–71].

In the past 12 month, we reviewed the literature and identified the need for an information-theoretic account of these expectancies and surprise signals. We designed an Ideal Observer model for novelty detection, which estimate the transition probabilities between stimuli based on a limited amount of recent observations. The only 'free parameter' of this model is thus the size of the window within which samples are considered. Using simulations, this simple model was validated on existing seminal data from the literature, as shown below:

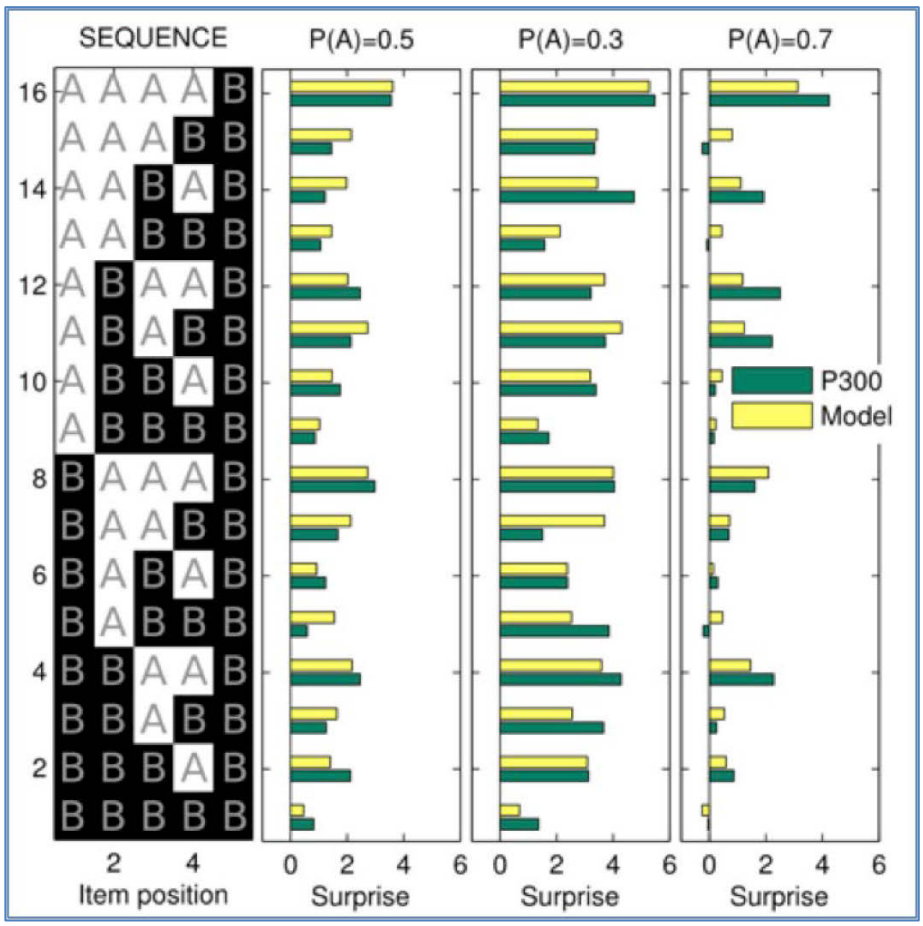


Figure 18: Surprise measured by P300 (data from Squires *et al.*, 1976) and model prediction.

Three long sequences of binary stimuli ('A' and 'B') were presented to subjects. These sequences were generated randomly, such that the average proportion of A was 50%, 30% or 70%. Within each sequence, the amplitude of the P300 waveform elicited by the stimulus B was averaged after splitting these responses according to the preceding pattern of 4 stimuli (the 16 possible patterns are listed line-wise on the left). Similarly, the 'surprise' of the model is shown in yellow (the surprise corresponds formally to -log(p), where p is the likelihood of observing B given the past observations). Note that numeric values were arbitrarily scaled so that the P300 and the model are matched on average. The model reproduces several features of the P300: it is impacted by the proportion of A, globally and within the local pattern, it is also impacted by the local pattern of alternation or repetition.

Another prediction of the model is that surprise signals should be impacted by the transition probability between stimuli, not only locally, but also globally. To test our model quantitatively against previous propositions, we planned to acquire new data with an extended version of the initial binary sequence protocol developed by Squires et al.[66].

In the initial protocol, transition probabilities were not biased globally and corresponded to the base rate of occurrence of each stimuli: the protocol had two conditions; block #1: p(A|B) = p(A|A) = 0.5; block #2: p(A|B) = p(A|A) = 0.3. The new protocol includes two additional blocks, in which the base rate is the same as in block #1, but transition probabilities are biased toward alternations in block #3: p(A|B) = p(B|A) = 0.3 or repetitions in block #4: p(A|B) = p(B|A) = 0.7.

7.1.5 Experiment 3. Non-Linguistic Nested structures (SP3_SKPI-16)

In Experiment 3, we plan to specifically test whether and how humans encode nested as opposed to serial regularities in sequences. To create non-trivial nested regularities outside the language domain, we designed a completely novel paradigm where a sequence of locations is presented visually, and sequence complexity is modulated by the number of (possibly nested) rules needed to encode (or compress) the sequence structure. For optimal comparability, all sequences use 8 locations on an octagon, but some sequences appear random (maximal complexity), some are compressible according to a nested rule of variable complexity, and some are only encodable as serial structures of variable length. We are comparing behaviour (anticipatory eye movements) and brain activity to those three types of sequences (random, nested or serial).

We have acquired adult eye-movement data, are starting data collection with adult human fMRI, and also plan (with non-HBP funding) to compare those data with similar data acquired in non-human primates. The outcome will reveal (a) whether non-linguistic sequences are encoded as serial or as nested structures (b) which areas are involved, and how their activity with complexity (c) whether human and non-human primates differ in this respect.

7.1.6 Plans for the next six months

The next six months will be devoted to:

- Experiment 1: Finish the acquisition of the experiment on mathematical syntax and work on the analyses of both experiments.
- Experiment 2: Acquire 2 pilots in the MEG (September – October), design and implement a pipeline for the analysis (September – January). Depending on the pilots' results, acquire a cohort of 20 subjects in the MEG (January-February)
- Experiment 3: Finish the acquisition of the fMRI experiment.
- Review: Start writing.

7.2 The Social Brain – Representing the Self in Relation to Others (T3.6.3)

7.2.1 Research goals

This Task will provide constraints for the emergence and shaping of the conscious mind, specifically to argue for the importance of other people in this process. The key scientific problem to tackle 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. The strategic experimental protocols will compare brain function during reactive, proactive and interactive states using simultaneous measurements of the brains of two interactors.

The group led by Riitta Hari (AALTO) will design and conduct neuroimaging experiments on humans to address these constraints. The experimental work will include fMRI and MEG

measurements of single subjects using simplified social stimuli (movies of encounters, persons, faces, eyes) with the goal to study brain function in as naturalistic conditions as possible. In addition, simultaneous measurements of two interacting subjects will be conducted. These dual measurements allow us to highlight brain processes supporting social interaction by seeking for commonalities (e.g. synchrony) in the signals of the two brains, without reference to the typically complex behavioural patterns of the dyad.

7.2.2 Main achievements (SP3_SKPI-17)

Riitta Hari's group (AALTO) reviewed literature on suggested experimental protocols for localisation of brain regions involved in social cognition. We have collected published stimulus sets and generated new stimuli to identify brain functions of interest for social cognition and for observed social interaction. The first version of the experimental protocol ("social localiser") is ready to be piloted using fMRI. In addition, new data from simultaneous MEG measurements of two interacting subjects have been collected from 16 subject-pairs, one of whom was measured at the Aalto University, Espoo, Finland, and the other at the same time at the Erasmus Hospital, Brussels, Belgium.

Figure 19: MEG-to-MEG experimental set-up

Simultaneous MEG measurements of interacting subject-pairs enable studies on real social interaction. Here the subjects were performing synchronised motor actions.

7.2.3 Plans for the next six months

During M13–M18 (October 2014 to April 2015), we will finalise the social localiser experiment, run the experiment on a group of subjects and perform the preliminary data analysis. In addition, the work on the acquisition and analysis of the new data from the dual imaging experiments will continue.

8. Scientific Coordination (WP3.7)

Stanislas Dehaene (CEA), SP3 leader and Thiên-Ly Pham (project manager, CEA) have coordinated SP3 activities. They have ensured that the Deliverables at Month 12 are on track: draft cognitive architectures and strategic experimental protocols.

8.1 Draft cognitive architectures: A special issue of a major peer-reviewed journal

One of the Milestones at Month 12 is a "draft cognitive architecture": a document that prefigures, in draft form, the delivery at the end of the ramp-up phase (month 30) of definitive **synthetic summaries of neuro-cognitive constraints on cognitive architectures**.

To maximise the scientific utility and quality of these documents and to ensure their peer review, Stanislas Dehaene has proposed that these Deliverables are ultimately published together as a **special issue of a major journal.** This is a highly innovative use of the Deliverables of an EU project. Each group of HBP scientists, focusing on a given cognitive function, will write its Deliverable in the form of a ~10-page opinion paper on its area of cognitive neuroscience. The whole, consisting of about 15 such reviews, will be a major contribution.

Stanislas Dehaene and Yadin Dudai (WIS) will be the editors of this special issue. Close discussions have been started with Katja Brose, editor of *Neuron,* who strongly supports the project. Together, the editors have defined the aim of the special issue: to survey a number of areas of cognitive neuroscience and, for each of them, to summarise the essential facts, laws or "benchmark" data (behavioural, neurophysiological, anatomical…) that any theorist should know before attempting to develop a model. They have decided that, in a number of cases, it might be useful to adjoin additional authors belonging to others Subprojects of the HBP, or even outside the HBP, in order to maximise the scientific expertise as well as geographic and gender diversity. They have also defined a canvass for each paper, which should include two boxes: "what we know" (simple summary statements of the main facts) and "what we need to know" (fundamental questions for future research and modelling).

The plan of the special issue is currently as follows:

- Introduction to special issue (1 page)
  - Stanislas Dehaene and Yadin Dudai
  - Introduction to the concept of "cognitive architectures", and the problem of bridging from neural populations to cognitive computations and behaviour. Selection of a small set of "Hilbert problems for cognitive neuroscience: landmark cognitive computations whose underlying mechanism remains unsolved.
- Visual Ignitions: Non-linear dynamics underlying the crossing of visual awareness thresholds in the human brain

  Rafi Malach, Clément Moutard, Stanislas Dehaene
- Visual Perception of Actions: Constraints from Neurophysiology and Computation
  - Martin Giese
- Attention and the modulation of inter-areal communication
  - Pascal Fries

- Cortical correlates of low-level perception: unimodal and multimodal sensory integration
  - Yves Frégnac and Brice Bathellier
- Neuroscientific constraints for computational approaches of self-consciousness
  - Olaf Blanke and Mel Slater
- A multi-level organisation for the sense of confidence in the brain
  - Florent Meyniel, Mariano Sigman, Zach Mainen
- Consolidation: Shaping and Reshaping Memory
  - Yadin Dudai, Avi Karni, Rony Paz, Jan Born
- The cognitive architecture of working memory
  - Lars Nyberg, Misha Tsodyks, Ed Vogel
- The cognitive architecture of spatial navigation: hippocampal and striatal contributions
  - Neil Burgess
- Varieties of sequence knowledge: From transition probabilities to symbolic rules and linguistic trees
  - Stanislas Dehaene, Florent Meyniel, Catherine Wacongne, Liping Wang, Christophe Pallier
- Centrality of social interaction in human brain function
  - Riitta Hari and Lauri Parkkonen
- Development of the human brain: cognitive and anatomical constraints
  - Ghislaine Dehaene-Lambertz and Elizabeth Spelke (to be confirmed)
- Understanding human cortical topography and functional specialisation
  - Katrin Amunts and Kalanit Grill-Spector (to be confirmed)

As planned, most of the authors have provided their drafts (month 12 Deliverable), which are annexed to this document (Annex C).

8.2 Strategic experimental protocols

A second Milestone at Month 12 is the delivery of strategic experimental paradigms ("localisers") for each of the major cognitive functions under study. After reviewing the literature and, in many cases, running new experiments, the Task leaders have provided a description of their best fMRI or MEG experiments, designed to parse their cognitive function of interest. All of these documents are provided in **Annex D**:

- Mapping individually-specific neuro-cognitive biases in the human brain (Rafi Malach group)
- Selective attention for modulation of inter-areal communication (Pascal Fries group)
- Neural correlates of bodily self-consciousness (Olaf Blanke group)
- A localiser for confidence (Mariano Sigman & Florent Meyniel)
- Localiser – Skills and habits (procedural memory) (Avi Karni group)

- Localiser for brain circuits that trigger consolidation of realistic episodes (Yadin Dudai group)
- Localiser – Working memory (Lars Nyberg group)
- Localiser for hippocampal spatial memory (Neil Burgess group)
- Localiser for language comprehension (Christophe Pallier group)
- Social localiser (Riitta Hari group)

The scientific coordinator, Stanislas Dehaene, will transmit these documents to Bertrand Thirion (Inria), in charge of the Individual Brain Charting project in SP2. Dehaene and Thirion are meeting regularly in order to decide which of these localisers will be run in the 10 individual subjects whose brain will be extensively mapped during the 10-year course of the HBP.

8.3 Measurement of progress (see Annex B)

The progress of work is measured using Key Performance Indicators as determined in the HBP Description of Work. The coordinator evaluates the progression of experimental data acquisition by examining how far research has progressed on the following scale:

- Definition of a strategic theory-relevant experiment, given the existing literature
- Experimental design
- Generation of stimuli
- Programming of the experiment
- Pilot data
- Acquisition of full data set
- Programming of data analyses
- Generation of definitive results
- Paper publication

The progression of simulation models (relevant only for a subset of the proposed cognitive architectures) is evaluated according to the following scale (see Tasks T3.1.2 led by Martin Giese and WP3.4 led by Neil Burgess):

- Definition of strategic data for modelling, given the existing literature
- Simulation design
- Pilot simulation data
- Generation of definitive results
- Comparison with strategic data
- Paper publication

Finally, progress in defining synthetic summaries of neuro-cognitive constraints on cognitive architectures is evaluated according to the number of syntheses and, for each of them, the following criteria:

- Isolation of relevant data
- Write up

- Synthesis and, if appropriate, corresponding data put in appropriate form on the HBP KnowledgeSpace
- Validation by one or more external referees

Expected times of delivery from each team have been defined.

8.4 Meetings and coordination

Meetings have been regularly organised and scientific discussions with WP and Task leaders are ongoing. On the 11th and 12th of June 2014, a joint SP3-SP4 workshop (First Theory-Cognitive workshop) took place at EITN (European Institute for Theoretical Neuroscience).

SP3 will organise several additional conferences:

- April 2015: SP3 Conference in Paris
- 29 June – 3 July 2015: Treilles meeting on metacognition (organised with our Partners from Champalimaud Foundation)
- September 2015: Conference on thalamocortical dynamics (SP3-SP4)

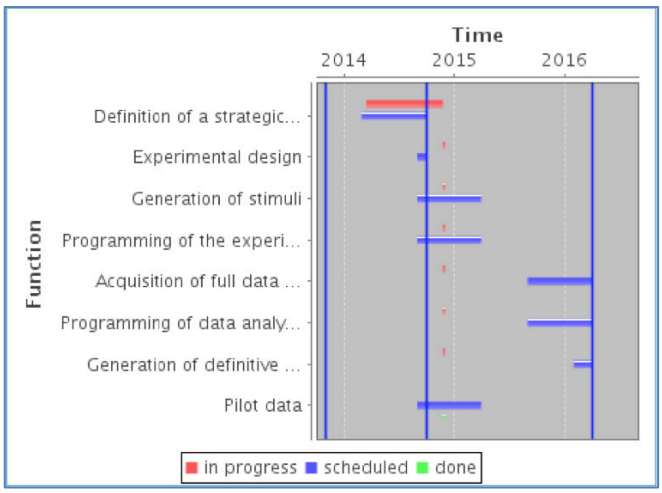
Annex A: Milestones

No.	Milestone Name	WP	Month Due	Month Achieved	See Page
M40	First-draft protocol for localising cognitive architecture using functional MRI defined.	3.1	6	6	
M45	First-draft protocol for localising cognitive architecture using functional MRI defined.	3.2	6	6	
M49	First-draft protocol for localising cognitive architecture using functional MRI defined.	3.3	6	6	
M53	First-draft protocol for localising cognitive architecture using functional MRI defined.	3.4	6	6	
M58	First-draft protocol for localising cognitive architecture using functional MRI defined.	3.5	6	6	
M62	First-draft protocol for localising cognitive architecture using functional MRI defined.	3.6	6	6	
M41	Community consensus on localisers; first datasets generated and localisers validated; first draft cognitive architectures.	3.1	12	12	
M46	Community consensus on localisers; first datasets generated and localisers validated; first draft cognitive architectures.	3.2	12	12	
M50	Community consensus on localisers; first datasets generated and localisers validated; first draft cognitive architectures.	3.3	12	12	
M54	Community consensus on localisers; first datasets generated and localisers validated; first draft cognitive architectures.	3.4	12	12	
M59	Community consensus on localisers; first datasets generated and localisers validated; first draft cognitive architectures.	3.5	12	12	
M63	Community consensus on localisers; first datasets generated and localisers validated; first draft cognitive architectures.	3.6	12	12	

Annex B: Scientific Key Performance Indicators (SKPIs)

SP3_SKPI-01 Study of the circuits involved in non-conscious and conscious mechanisms of visual recognition. Responsible: rafi.malach@gmail.com

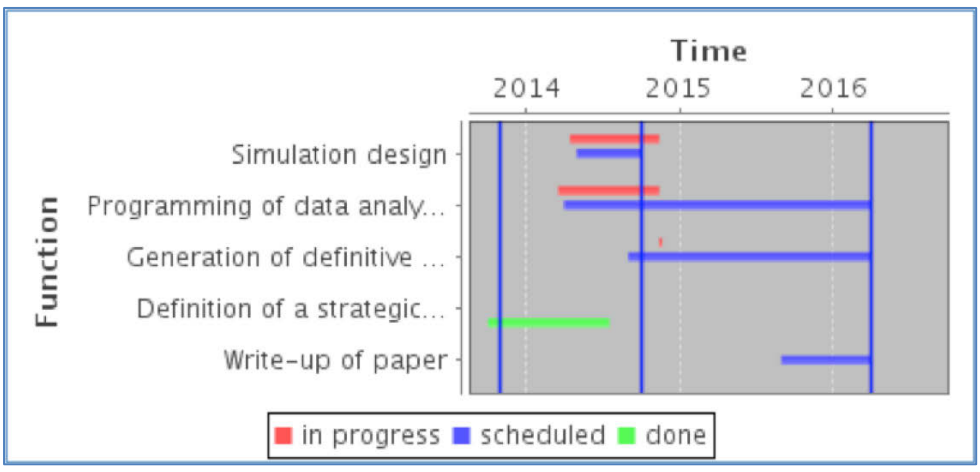
- Definition of a strategic question, given the existing literature. Planned: 2014/02/28 - 2014/09/30
- Experimental design. Planned: 2014/08/31 - 2014/09/30
- Generation of stimuli. Planned: 2014/08/31 - 2015/03/31
- Programming of the experiment. Planned: 2014/08/31 - 2015/03/31
- Pilot data. Planned: 2014/08/31 - 2015/03/31
- Acquisition of full data set. Planned: 2015/08/31 - 2016/03/31
- Programming of data analyses. Planned: 2015/08/31 - 2016/03/31
- Generation of definitive results. Planned: 2016/01/31 - 2016/03/31



SP3_SKPI-02 Understanding the circuits linking perceptions to actions

Responsible: martin.giese@uni-tuebingen.de

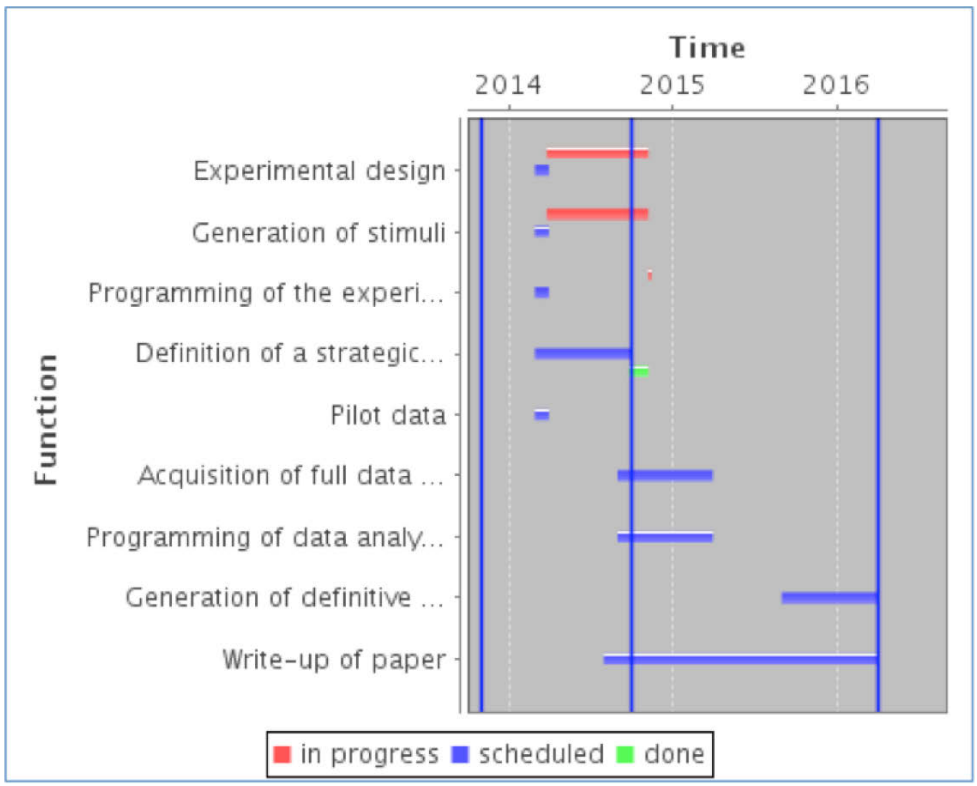
- Definition of a strategic question, given the existing literature. Planned: 2014/03/31 - 2014/03/31
- Simulation design. Planned: 2014/04/30 - 2014/09/30
- Programming of data analyses. Planned: 2014/03/31 - 2016/03/31
- Generation of definitive results. Planned: 2014/08/31 - 2016/03/31
- Write-up of paper. Planned: 2015/08/31 - 2016/03/31



SP3_SKPI-03 Understanding how body perception becomes a reference point for the sense of self

Responsible: olaf.blanke@epfl.ch

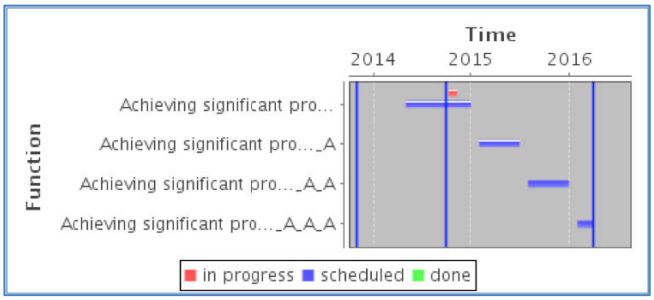
- Definition of a strategic question, given the existing literature. Planned: 2014/02/28 - 2014/09/30
- Experimental design. Planned: 2014/02/28 - 2014/03/31
- Generation of stimuli. Planned: 2014/02/28 - 2014/03/31
- Programming of the experiment. Planned: 2014/02/28 - 2014/03/31
- Pilot data. Planned: 2014/02/28 - 2014/03/31
- Acquisition of full data set. Planned: 2014/08/31 - 2015/03/31
- Programming of data analyses. Planned: 2014/08/31 - 2015/03/31
- Generation of definitive results. Planned: 2015/08/31 - 2016/03/31
- Write-up of paper. Planned: 2014/07/31 - 2016/03/31



SP3_SKPI-19 Multi-scale data analysis and multi-scale transfer modelling

Responsible: p.deweerd@maastrichtuniversity.nl

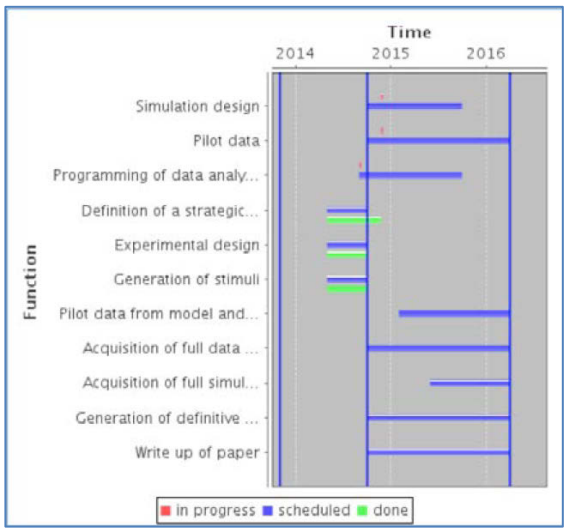
- Achieving significant progress towards establishing the key parameters for an unstructured (P)ING ne. Planned: 2014/04/30 - 2014/12/31
- Achieving significant progress towards the use of physiologically validated model parameters to crea. Planned: 2015/01/31 - 2015/06/30
- Achieving significant progress towards the development of a structured, expanded (topographic) model. Planned: 2015/07/31 - 2015/12/31
- Achieving significant progress towards using the expanded model to study differences/commonalities b. Planned: 2016/01/31 - 2016/03/31



SP3_SKPI-21 Development and validation of brain network models constrained by realistic physiological phase lags

Responsible: viktor.jirsa@univ-amu.fr

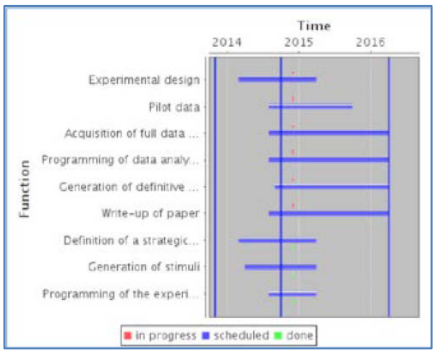
- Definition of a strategic question, given the existing literature. Planned: 2014/04/30 - 2014/09/30
- Experimental design. Planned: 2014/04/30 - 2014/09/30
- Simulation design. Planned: 2014/09/30 - 2015/09/30
- Generation of stimuli. Planned: 2014/04/30 - 2014/09/30
- Pilot data. Planned: 2014/09/30 - 2016/03/31
- Pilot data from model and simulations. Planned: 2015/01/31 - 2016/03/31
- Acquisition of full data set. Planned: 2014/09/30 - 2016/03/31
- Acquisition of full simulation data set. Planned: 2015/05/31 - 2016/03/31
- Programming of data analyses. Planned: 2014/08/31 - 2015/09/30
- Generation of definitive results. Planned: 2014/09/30 - 2016/03/31
- Write up of paper. Planned: 2014/09/30 - 2016/03/31



SP3_SKPI-04 Mapping and understanding the neuronal circuits involved in decision making, confidence and error co

Responsible: mariuchu@gmail.com

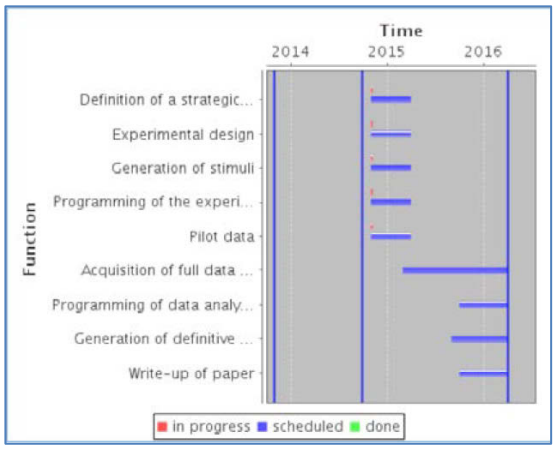
- Definition of a strategic question, given the existing literature. Planned: 2014/02/28 - 2015/03/31
- Experimental design. Planned: 2014/02/28 - 2015/03/31
- Generation of stimuli. Planned: 2014/03/31 - 2015/03/31
- Programming of the experiment. Planned: 2014/07/31 - 2015/03/31
- Pilot data. Planned: 2014/07/31 - 2015/09/30
- Acquisition of full data set. Planned: 2014/07/31 - 2016/03/31
- Programming of data analyses. Planned: 2014/07/31 - 2016/03/31
- Generation of definitive results. Planned: 2014/08/31 - 2016/03/31
- Write-up of paper. Planned: 2014/07/31 - 2016/03/31



SP3_SKPI-05 Mapping and understanding the neuronal circuits involved in motivation, emotion and reward

Responsible: mathias.pessiglione@gmail.com

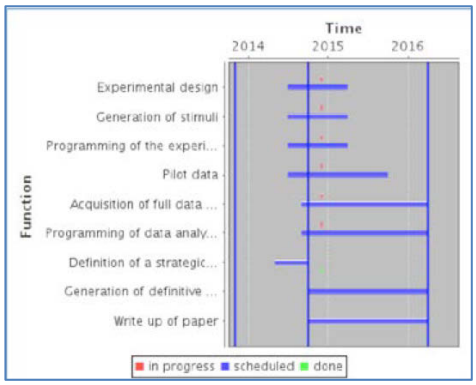
- Definition of a strategic question, given the existing literature. Planned: 2014/10/31 - 2015/03/31
- Experimental design. Planned: 2014/10/31 - 2015/03/31
- Generation of stimuli. Planned: 2014/10/31 - 2015/03/31
- Programming of the experiment. Planned: 2014/10/31 - 2015/03/31
- Pilot data. Planned: 2014/10/31 - 2015/03/31
- Acquisition of full data set. Planned: 2015/02/28 - 2016/03/31
- Programming of data analyses. Planned: 2015/09/30 - 2016/03/31
- Generation of definitive results. Planned: 2015/08/31 - 2016/03/31
- Write-up of paper. Planned: 2015/09/30 - 2016/03/31



SP3_SKPI-20 Dissecting the brainstem modulation of cortical decision computations

Responsible: t.h.donner@uva.nl

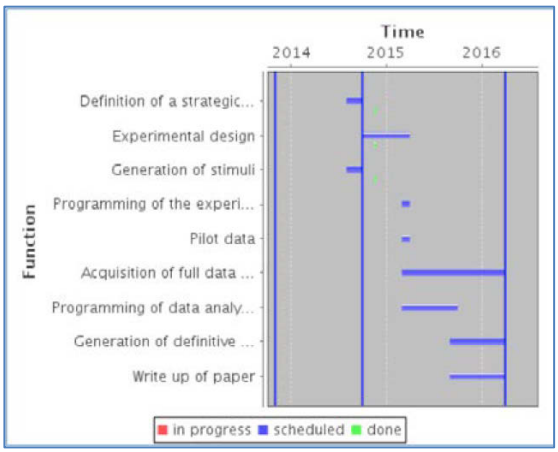
- Definition of a strategic question, given the existing literature. Planned: 2014/04/30 - 2014/09/30
- Experimental design. Planned: 2014/06/30 - 2015/03/31
- Generation of stimuli. Planned: 2014/06/30 - 2015/03/31
- Programming of the experiment. Planned: 2014/06/30 - 2015/03/31
- Pilot data. Planned: 2014/06/30 - 2015/09/30
- Acquisition of full data set. Planned: 2014/08/31 - 2016/03/31
- Programming of data analyses. Planned: 2014/08/31 - 2016/03/31
- Generation of definitive results. Planned: 2014/09/30 - 2016/03/31
- Write up of paper. Planned: 2014/09/30 - 2016/03/31



SP3_SKPI-18 Characterise Multi-scale brain architecture of decision related motivational states and values

Responsible: tomergazit@gmail.com

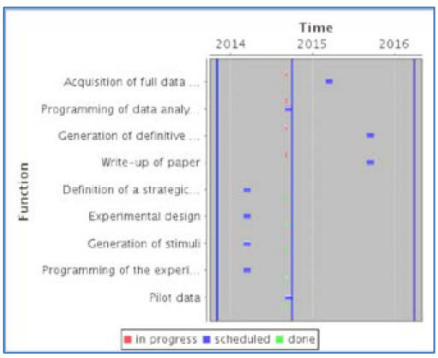
- Definition of a strategic question, given the existing literature. Planned: 2014/07/31 - 2014/09/30
- Experimental design. Planned: 2014/09/30 - 2015/03/31
- Generation of stimuli. Planned: 2014/07/31 - 2014/09/30
- Programming of the experiment. Planned: 2015/02/28 - 2015/03/31
- Pilot data. Planned: 2015/02/28 - 2015/03/31
- Acquisition of full data set. Planned: 2015/02/28 - 2016/03/31
- Programming of data analyses. Planned: 2015/02/28 - 2015/09/30
- Generation of definitive results. Planned: 2015/08/31 - 2016/03/31
- Write up of paper. Planned: 2015/08/31 - 2016/03/31



SP3_SKPI-06 Brain signatures of motor skill consolidation in young healthy human adults (fMRI study)

Responsible: avik@construct.haifa.ac.il

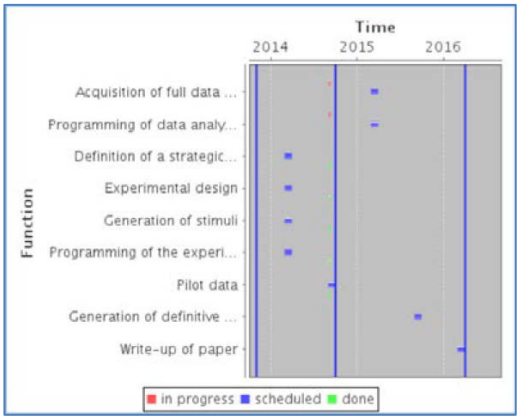
- Definition of a strategic question, given the existing literature. Planned: 2014/02/28 - 2014/03/31
- Experimental design. Planned: 2014/02/28 - 2014/03/31
- Generation of stimuli. Planned: 2014/02/28 - 2014/03/31
- Programming of the experiment. Planned: 2014/02/28 - 2014/03/31
- Pilot data. Planned: 2014/08/31 - 2014/09/30
- Acquisition of full data set. Planned: 2015/02/28 - 2015/03/31
- Programming of data analyses. Planned: 2014/08/31 - 2014/09/30
- Generation of definitive results. Planned: 2015/08/31 - 2015/09/30
- Write-up of paper. Planned: 2015/08/31 - 2015/09/30



SP3_SKPI-07 Differential effects of observation and actual movement performance on procedural memory consolidation

Responsible: avik@construct.haifa.ac.il

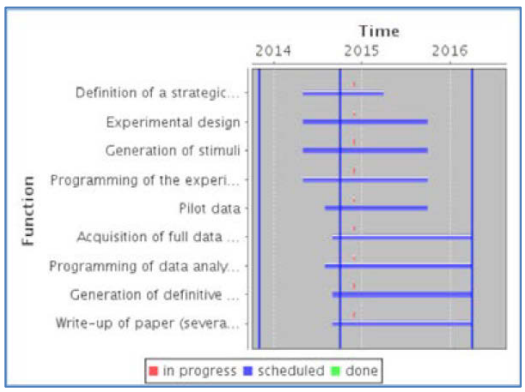
- Definition of a strategic question, given the existing literature. Planned: 2014/02/28 - 2014/03/31
- Experimental design. Planned: 2014/02/28 - 2014/03/31
- Generation of stimuli. Planned: 2014/02/28 - 2014/03/31
- Programming of the experiment. Planned: 2014/02/28 - 2014/03/31
- Pilot data. Planned: 2014/08/31 - 2014/09/30
- Acquisition of full data set. Planned: 2015/02/28 - 2015/03/31
- Programming of data analyses. Planned: 2015/02/28 - 2015/03/31
- Generation of definitive results. Planned: 2015/08/31 - 2015/09/30
- Write-up of paper. Planned: 2016/02/29 - 2016/03/31



SP3_SKPI-08 Memory for facts and events

Responsible: yadin.dudai@weizmann.ac.i

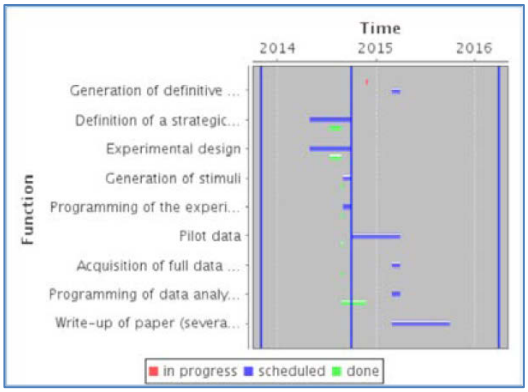
- Definition of a strategic question, given the existing literature. Planned: 2014/04/30 - 2015/03/31
- Experimental design. Planned: 2014/04/30 - 2015/09/30
- Generation of stimuli. Planned: 2014/04/30 - 2015/09/30
- Programming of the experiment. Planned: 2014/04/30 - 2015/09/30
- Pilot data. Planned: 2014/07/31 - 2015/09/30
- Acquisition of full data set. Planned: 2014/08/31 - 2016/03/31
- Programming of data analyses. Planned: 2014/07/31 - 2016/03/31
- Generation of definitive results. Planned: 2014/08/31 - 2016/03/31
- Write-up of paper (several papers written along the line, not only one paper). Planned: 2014/08/31 - 2016/03/31



SP3_SKPI-09 Working memory

Responsible: lars.nyberg@physiol.umu.se

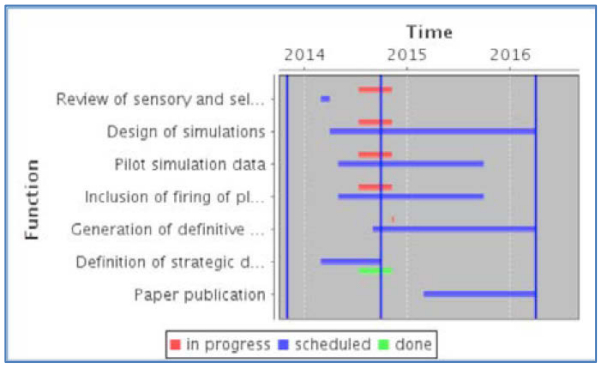
- Definition of a strategic question, given the existing literature. Planned: 2014/04/30 - 2014/09/30
- Experimental design. Planned: 2014/04/30 - 2014/09/30
- Generation of stimuli. Planned: 2014/08/31 - 2014/09/30
- Programming of the experiment. Planned: 2014/08/31 - 2014/09/30
- Pilot data. Planned: 2014/09/30 - 2015/03/31
- Acquisition of full data set. Planned: 2015/02/28 - 2015/03/31
- Programming of data analyses. Planned: 2015/02/28 - 2015/03/31
- Generation of definitive results. Planned: 2015/02/28 - 2015/03/31
- Write-up of paper (several papers written along the line, not only one paper). Planned: 2015/02/28 - 2015/09/30



SP3_SKPI-10 Model of hippocampal navigation

Responsible: n.burgess@ucl.ac.uk

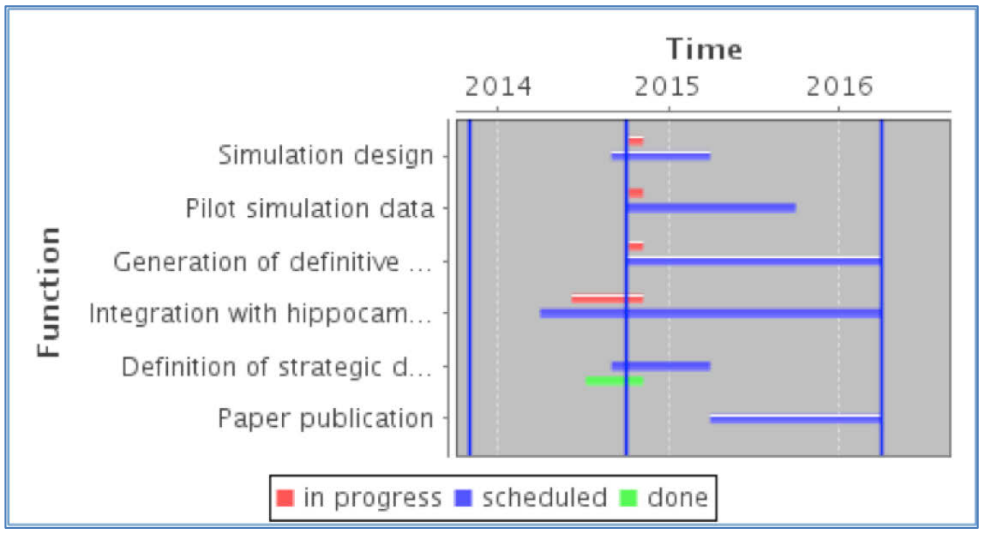
- Review of sensory and self-motion information in self-localisation and place and grid cell firing. Planned: 2014/02/28 - 2014/03/31
- Definition of strategic data for modelling based on existing literature. Planned: 2014/02/28 - 2014/09/30
- Design of simulations. Planned: 2014/03/31 - 2016/03/31
- Pilot simulation data. Planned: 2014/04/30 - 2015/09/30
- Inclusion of firing of place, head direction and grid cells. Planned: 2014/04/30 - 2015/09/30
- Generation of definitive results. Planned: 2014/08/31 - 2016/03/31
- Paper publication. Planned: 2015/02/28 - 2016/03/31



SP3_SKPI-11 Model of striatal navigation

Responsible: n.burgess@ucl.ac.uk

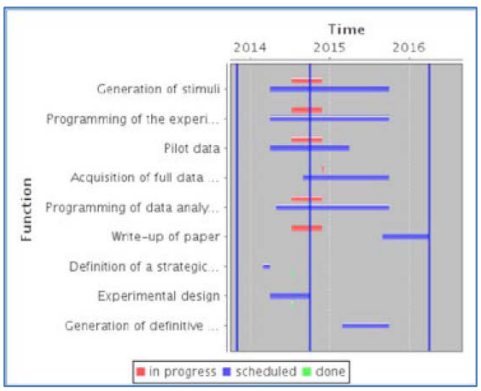
- Definition of strategic data for modelling, given the existing literature. Planned: 2014/08/31 - 2015/03/31
- Simulation design. Planned: 2014/08/31 - 2015/03/31
- Pilot simulation data. Planned: 2014/09/30 - 2015/09/30
- Generation of definitive results. Planned: 2014/09/30 - 2016/03/31
- Integration with hippocampal navigation model. Planned: 2014/03/31 - 2016/03/31
- Paper publication. Planned: 2015/03/31 - 2016/03/31



SP3_SKPI-12 Neural correlates of unimodal perception and self-organisation of internal knowledge in mammalian primary cortical areas

Responsible: fregnac@unic.cnrs-gif.fr

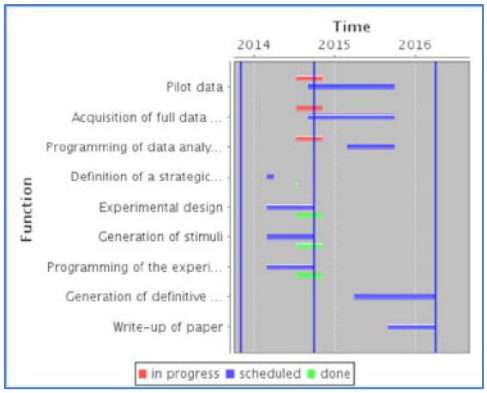
- Definition of a strategic question, given the existing literature. Planned: 2014/02/28 - 2014/03/31
- Experimental design. Planned: 2014/03/31 - 2014/09/30
- Generation of stimuli. Planned: 2014/03/31 - 2015/09/30
- Programming of the experiment. Planned: 2014/03/31 - 2015/09/30
- Pilot data. Planned: 2014/03/31 - 2015/03/31
- Acquisition of full data set. Planned: 2014/08/31 - 2015/09/30
- Programming of data analyses. Planned: 2014/04/30 - 2015/09/30
- Generation of definitive results. Planned: 2015/02/28 - 2015/09/30
- Write-up of paper. Planned: 2015/08/31 - 2016/03/31



SP3_SKPI-13 Neural correlates of unimodal and multi-modal perception in mammalian primary sensory areas

Responsible: brice.bathellier@unic.cnrs-gif.fr

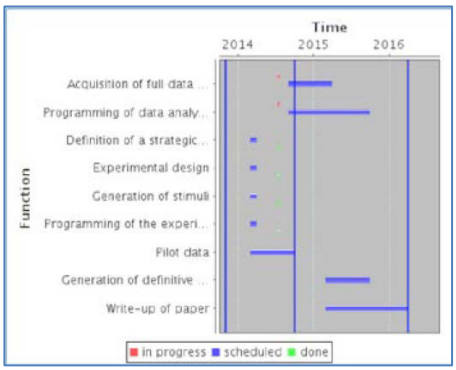
- Definition of a strategic question, given the existing literature. Planned: 2014/02/28 - 2014/03/31
- Experimental design. Planned: 2014/02/28 - 2014/09/30
- Generation of stimuli. Planned: 2014/02/28 - 2014/09/30
- Programming of the experiment. Planned: 2014/02/28 - 2014/09/30
- Pilot data. Planned: 2014/08/31 - 2015/09/30
- Acquisition of full data set. Planned: 2014/08/31 - 2015/09/30
- Programming of data analyses. Planned: 2015/02/28 - 2015/09/30
- Generation of definitive results. Planned: 2015/03/31 - 2016/03/31
- Write-up of paper. Planned: 2015/08/31 - 2016/03/31



SP3_SKPI-14 Syntactic Complexity of sentences: the role of empty categories

Responsible: christophe.pallier@polytechnique.org

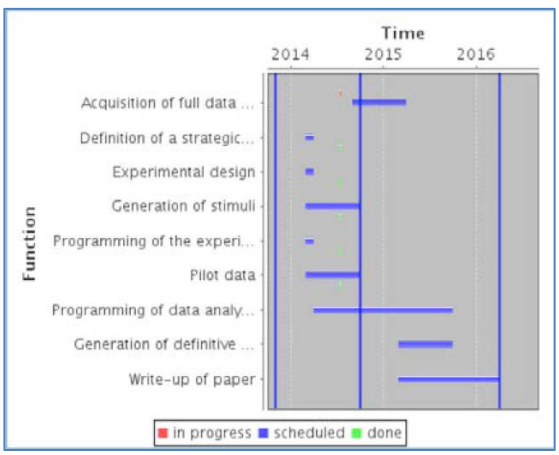
- Definition of a strategic question, given the existing literature. Planned: 2014/02/28 - 2014/03/31
- Experimental design. Planned: 2014/02/28 - 2014/03/31
- Generation of stimuli. Planned: 2014/02/28 - 2014/03/31
- Programming of the experiment. Planned: 2014/02/28 - 2014/03/31
- Pilot data. Planned: 2014/02/28 - 2014/09/30
- Acquisition of full data set. Planned: 2014/08/31 - 2015/03/31
- Programming of data analyses. Planned: 2014/08/31 - 2015/09/30
- Generation of definitive results. Planned: 2015/02/28 - 2015/09/30
- Write-up of paper. Planned: 2015/02/28 - 2016/03/31



SP3_SKPI-15 Bayesian Modelling of Expectation Effects in Sequences

Responsible: christophe.pallier@polytechnique.org

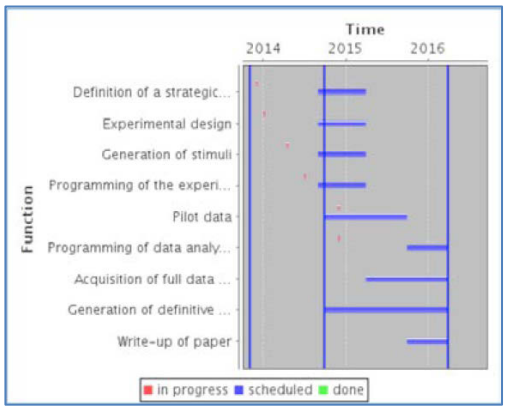
- Definition of a strategic question, given the existing literature. Planned: 2014/02/28 - 2014/03/31
- Experimental design. Planned: 2014/02/28 - 2015/09/30
- Generation of stimuli. Planned: 2014/02/28 - 2014/09/30
- Programming of the experiment. Planned: 2014/08/31 - 2014/09/30
- Pilot data. Planned: 2014/08/31 - 2014/09/30
- Acquisition of full data set. Planned: 2015/02/28 - 2015/09/30
- Programming of data analyses. Planned: 2015/02/28 - 2015/09/30
- Generation of definitive results. Planned: 2015/04/30 - 2015/09/30
- Write-up of paper. Planned: 2015/02/28 - 2016/03/31



SP3_SKPI-16 Geometrical artificial language with nested and non-nested sequences

Responsible: christophe.pallier@polytechnique.org

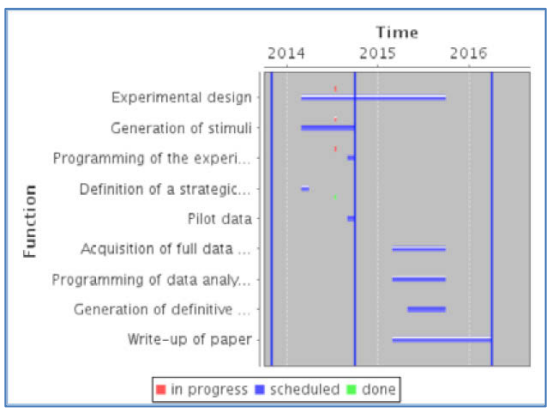
- Definition of a strategic question, given the existing literature. Planned: 2014/02/28 - 2014/03/31
- Experimental design. Planned: 2014/02/28 - 2014/03/31
- Generation of stimuli. Planned: 2014/02/28 - 2014/09/30
- Programming of the experiment. Planned: 2014/02/28 - 2014/03/31
- Pilot data. Planned: 2014/02/28 - 2014/09/30
- Acquisition of full data set. Planned: 2014/08/31 - 2015/03/31
- Programming of data analyses. Planned: 2014/03/31 - 2015/09/30
- Generation of definitive results. Planned: 2015/02/28 - 2015/09/30
- Write-up of paper. Planned: 2015/02/28 - 2016/03/31



SP3_SKPI-17 The social brain-representing the self in relation to others

Responsible: hari@neuro.hut.fi

- Definition of a strategic question, given the existing literature. Planned: 2014/08/31 - 2015/03/31
- Experimental design. Planned: 2014/08/31 - 2015/03/31
- Generation of stimuli. Planned: 2014/08/31 - 2015/03/31
- Programming of the experiment. Planned: 2014/08/31 - 2015/03/31
- Pilot data. Planned: 2014/09/30 - 2015/09/30
- Acquisition of full data set. Planned: 2015/03/31 - 2016/03/31
- Programming of data analyses. Planned: 2015/09/30 - 2016/03/31
- Generation of definitive results. Planned: 2014/09/30 - 2016/03/31
- Write-up of paper. Planned: 2015/09/30 - 2016/03/31



Annex C: Drafts of review papers for a special issue of Neuron on "cognitive architectures"

Visual Ignitions: Non-linear dynamics underlying the crossing of visual awareness thresholds in the human brain

Authors: Rafi Malach (WIS), Clément Moutard (CEA), Stanislas Dehaene (CEA)

Despite its introspectively direct and vivid character- the emergence of a visual image in the mind of a human observer has presented one of the most challenging and yet fascinating topics of research. Given that it has been by now well established that substantial level of visual processing can be accomplished in the absence of awareness- a critical question, on which we would like to focus in the present perspective, is what are the neuronal dynamics- that enable the crossing of visual processes from the unconscious, non-reportable domain to full awareness in which report, recall and meta-cognition become readily available to the individual.

Here we would like to propose that the critical dynamics can be metaphorically viewed as an "ignition" phenomenon- i.e. a highly non-linear process where crossing the awareness threshold leads to a rapid, self-amplifying increase in neuronal activity, that is can be sustained far beyond the original stimulus duration.

Theoretically, non-linear dynamics has a long history- beginning with Hebb's idea of a neuronal assembly, followed by point- attractor dynamics, and more recently global Ignition events in the context of global workspace models.

Importantly, behavioural studies, using paradigms such as backward masking, now converge in pointing to such threshold non-linearities in visual recognition as well.

The neuronal non-linear "ignition" processes can be revealed in BOLD-fMRI (Grill-Spector et al., 2000)[72], EEG recordings (Del Cul et al., 2007)[73], in direct intra-cranial recordings of mass neuronal activity (ECoG) (Fisch et al., 2009)[74] and even in rare single neuron recordings obtained in medial-temporal cortical structures (Quiroga et al., 2008)[75].

The neuronal mechanism driving such ignition phenomena are not known at present- but a likely anatomical substrate is readily available in the massive, and likely scale free, networks of intrinsic cortical connections (Amir et al., 1993)[76]. Such networks naturally form massive feed-back and recurrent loops. Thus, under appropriate levels of network excitability and specific stimulations such recurrent networks could cross the Ignition threshold. This threshold crossing, as in any Ignition phenomena, leads to self-driven increase in neuronal firing fed by positive-feedback loops among the relevant neurons contributing to the Ignited assembly. The link to conscious recognition is then straightforward- the Ignition threshold reflected in the neuronal domain is the basis of the recognition threshold revealed in visual processing.

The Ignition dynamics serves the important function of amplifying and stabilising a visual recognition even in the face of noisy and rapidly changing inputs.

The well-defined awareness threshold provides a ready explanation to the apparent conundrum posed by the recent discovery of on-going spontaneous fluctuations in neuronal activity that occur in the visual system in the complete absence of sensory. This activity is perplexing since in BOLD fMRI it appears to be of a similar magnitude to visually driven activations that are clearly associated with visual awareness (Nir et al., 2006)[77]. However, a number of studies have now argued that such spontaneous activations go below the awareness thresholds. Thus, visual imagery cannot account for such fluctuations, participants are not aware that they exist even while they in fact linked to their slow eye

movements, and perhaps most convincingly, participants deny any knowledge of these activation even when they receive clear feedback about their occurrence. The conundrum is readily resolved when considering that these fluctuations simply occur below the Ignition threshold. Although in BOLD imaging the amplitude of the spontaneous fluctuations appear to be of similar magnitude to that of sensory stimuli, it is important to note that these fluctuations are of ultra-slow dynamics. Such slow events are expected to be artificially "amplified" by the low-pass temporal filtering of the BOLD hemodynamics response (Nir et al., 2008)[78]. Thus, under the "ignition" hypothesis- it is predicted that spontaneous fluctuations should always remain below the Ignition threshold. If due to changes in excitability state or even due to random variations, a spontaneous fluctuation happens to cross the awareness threshold, than according to the Ignition hypothesis, an Ignition will occur, and the participant will experience a visual image whose contents correspond to the functional selectivity of the "ignited" assembly.

Interestingly, recent research, attempting to examine the possible link between spontaneous neuronal activity fluctuations and reportable subjective events appears to be intriguingly compatible with such hypothesis. In the experiments, spontaneous modulations in firing rate of single neurons were recorded in patients implanted with depth electrodes for clinical diagnostic purposes, while they were engaged in a free recall paradigm. Importantly, concurrent with the neuronal recordings, the patients were asked to verbally report the stream of thoughts images that entered into their awareness. Remarkably, as predicted by the Ignition hypothesis, the neuronal firing underwent continuous fluctuations- however, as long as the amplitude of these firing rate fluctuations were low- there was no relationship between the patients' reports and the content selectivity of the neurons. In contrast, whenever a fluctuation attained exceptionally high amplitude- the patients' report aligned with the previously recorded selectivity of the neuron (Gelbard-Sagiv et al., 2008)[79]. A similar link between high amplitude fluctuations and voluntary decisions has been proposed in a libet-like paradigm (Schurger et al., 2012)[80]. Recently, an ignition-type dynamics has been implicated in a correlation found between spontaneous BOLD fluctuations and pupillary diameter. Given the established link between pupil diameter and attention- this observation is compatible with the proposed connection neuronal ignitions and crossing the awareness threshold.

While the concept of Ignition appears to be compatible with a number of experimental observations related to human conscious recognition- it should be emphasised that clearly the phenomena is far more complex than a simple all-or none "spike" like ignition of activity- and many fascinating questions still remain to be explored. Below we outline a few.

First, the notion of a binary, stereotypical ignition of fixed amplitude is at odds with gradual amplitude changes that can be consistently observed during conscious recognition. An important example is the adaptation phenomena- in which repeated presentations of the same stimulus leads to a gradual reduction in the amplitude of activation. This activity change apparently occurs in the absence of a discernable change in recognition ability or visual awareness. Along the same lines, the temporal evolvement of the neuronal responses associated with the Ignition phenomena is quite complex. Rather than the stereotypical rapid climbing towards the maximum activation level and a sustained response at this maximum level- the typical neuronal response profile shows a high amplitude transient response, followed by a slowly declining sustained signal. Clearly, the simple notion of Ignition as the sole correlate of visual awareness will need to be modified to accommodate such more "analogue" phenomena.

Another question concerns the process that leads to the turning off of the Ignited activity- in a typical visual response- the duration of the neuronal activity (even in response to very brief, e.g. 20 msec stimulus) lasts for about 400 msec- and then starts declining. It is

currently unknown what is the mechanism that underlies this self-termination. A plausible candidate may be synaptic fatigue (which has also been implicated in adaptation) but this possibility has not been experimentally verified yet.

A complementary question concerns the issue of change detection- a clear advantage of the Ignition phenomenon- common to all such threshold non-linearities is their relative immunity to noise. Once an Ignition is initiated, its activity is self-sustained and it is therefore insensitive to small changes in input strength. However, such stability seems at odds with the excellent ability of the visual system to detect rapid transients. A particularly fascinating illustration of the complexity of this issue is the case of the momentary luminance changes associated with spontaneous blinks. While it may appear that such transient changes may be too fast for the visual system to detect- it turns out that participants can easily detect such changes when they are induced artificially, in the absence of blinks- yet are completely oblivious to their occurrence in the context of spontaneous blinks. It may appear that such selective sensitivity to transient darkening may be at odds with the all-or none nature of the ignition phenomenon. This issue merits further investigation.

It is important to acknowledge the methodological constraints in our ability to detect neuronal ignitions when they occur. This is particularly challenging in the case of human recognition- since the more direct means of detecting such ignitions- i.e. recording the firing rates of isolated neurons is extremely rare and confined to very specific clinical situations. Less detailed measurements, such as intra-cranial recordings of mass potentials (ECOG) and BOLD fMRI responses- are more common, however depend on the fact that the assembly participating in an ignition is densely confined to local points in the cortex. Sparse assemblies, in which only a small minority of neurons participate in the assembly in each cortical site may produce signals that are too weak to be detected by ECOG or fMRI. Fortunately, due to the organisation of the cortex into functionally selective columns, and the particular density of local reverberatory connections, there are many instances in which cortical Ignitions can be detected locally by examining the signal amplitude in local cortical sites. However, for the sparser representations, it may be that more sensitive signal recordings, such as multi-variate pattern analysis may be a better means to detect the emergence of assembly Ignitions.

Finally, from a somewhat philosophical perspective- it is important to note that Ignitions in general fall under the category of emergence- i.e. a phenomenon that cannot be decomposed into its elementary constituents. Thus, the non-linear nature of the neuronal ignition and its dependence on rapid, asynchronous and recurrent transfer of information- makes it is difficult to see how it could be possible to recreate such phenomena through a digitised simulation as detailed as that may be. There is a growing realisation of the importance of integrated information in generating the uniquely unitary conscious awareness event. In such a framework it is tempting to conjecture that the gestalt-like fusion of individual neurons' information through the nonlinear ignition process may endow the visual assembly with an integrative ingredient essential for the emergence of visual awareness.

List of essential facts or decisive statements

1) A critical element in determining the crossing of conscious awareness threshold is a non-linear "ignition" of neuronal activity. According to this hypothesis the neuronal implementation of the behavioural threshold of conscious awareness is the activity threshold at which such ignitions occur.

2) Behavioural evidence using methods such as backward masking is compatible with the notion of a consciousness threshold

3) Ignition phenomena have been associated in the human cortex in relationship with the consciousness threshold in all modalities- EEG, BOLD-fMRI, Intra-cranial ECOG recordings and single unit recordings.
4) The Ignition dynamics is compatible with a long history of theoretical modelling work- from the Hebbian concept of cell assembly, through point attractor dynamics to recent global Ignition models.
5) The concept of awareness threshold provides a ready explanation for the relationship between spontaneous fluctuations and cognitive awareness.
6) Open questions that need addressing include the gradual aspects of visual responses- e.g. during adaptation. Similarly, the mechanism of self-termination of the ignitions dynamic is not elucidated yet.
7) Ignition, being a highly non-linear phenomena, likely driven by asynchronous reverberatory signalling, is an emergent phenomena- and hence likely cannot be broken down to elementary components- e.g. in a digital simulation.

Visual Perception of Actions: Constraints from Neurophysiology and Computation

Authors: Martin Giese (EKUT)

Introduction

Perception of action and analysis of abstract action-related stimuli has been a central topic in cognitive neuroscience

Yet most underlying neuro-computational mechanisms remain unclear

Huge number of speculative conceptual theories; in addition, a number of computational theories, where often the detailed relationship with the behaviour of real neurons remains relatively vague

Goal of the paper: review the key results from physiology and discuss what they might imply in terms of constraints for computational mechanism that can be implemented with real neurons. Discussion of the existing models in the context of these constraints.

Key results from electrophysiology

Discussion why electrophysiological recordings are essential in order to determine computational mechanisms at the detailed level, and why this cannot be done with fMRI and EEG alone.

Focus on original physiological studies and implemented models (few secondary literature and nom extensive treatment of conceptual theories)

Short discussion of essential electrophysiological results with implication for computations from the literature and own experimental work:

- Visual selectivity for goal-directed actions in different regions (STS, parietal, F5); relationship between visual and motor properties
- Basic computational properties (position invariance, view dependence, sequence selectivity, selectivity for 'mimicked' actions)
- Hierarchical encoding of sequences; interpretation as 'goal' or 'intention encoding'
- Adaptation properties of action-selective neurons and relevance for fMRI
- Non-trivial generalisation properties (multi modality, robustness to occlusions, tuning for interrupted actions; generalisation to visual causality stimuli, neural signatures of encoding of semantic relationships)
- Additional properties: encoding of body scheme; encoding of own vs. other's actions; encoding of operational space; encoding of amounts of reward; relationship to decision

Theoretical models

- Focus on models with real quantitative implementations; beyond this many conceptual theories (cite other reviews)
- Classical neural network models (Arbib, Boniauto)
- Controller-based models (Oztop, Wolpert, Demiris)
- Bayesian models (Friston, Kilner)
- Example-based recognition models (Giese, Lappe, Kornprobst, Serre, Metta, Butz)
- Neural field models (Erlhagen, Billard)

- Physiologically-inspired models (Giese, Chersi)
- Implementations of learning

Proposal: neural mass models as intermediate level of description

- Definition of neural mass model; relationship with single cell; interpretation in terms of spiking neuron models and sampling; methods to analyse dynamics of relevant self-organised solutions; progress in the theory of neural fields
- Discussion of different relevant computational mechanisms that can be treated in this framework: sequence representation, decision, cue fusion, parameter estimation, coordinate transformation (Deneve), learning of coupled maps
- Similarities and difference with probabilistic models; limitations of present understanding of match between the two approaches; question in which sense neurons might approximate probabilistic inference

Open questions and future challenges

- More detailed localisation of individual computational steps; example: work by Orban et al. trying to establish fine-grained homologies between human and monkey cortex using fMRI
- More detailed understanding at the level of cell types and layers; data missing; new animal models required (rodents, marmosets)
- Exact theoretical understanding of relationship between neurodynamic approaches and statistical / Bayesian approaches
- Models that integrate relevant computational functions (e.g. sequence coding + cue fusion + memory + coordinate transformations) within a coherent neural architecture´
- Role of subcortical structures (cerebellum, basal ganglia)
- Exact definition of 'action semantics' in nonverbal cortical representations; how is such semantics encoded at the neural level
- Role of oscillations and synchroneous spiking patterns; relationship to relevant phenomena from human EEG studies
- Neural mechanisms for the disambiguation of the encoding of own body and actions, and the observed ones by others; relationship of such mechanisms with social context and attention

Box 1: Neural structures involved in the visual processing of observed actions:

Overview picture of monkey cortex, integrating recent results on cortical and sub-cortical structures neural structures involved in visual action processing.

Box 2: Example for a direct link between physiology and computation: physiologically-inspired neural model for the recognition of goal-directed actions

Sketch of Fleischer et al. 2013 model and sketch of additional modules that might be necessary for a more complete account of visual action processing.

Attention and the modulation of inter-areal communication

Authors: Chris Lewis (ESI), Pascal Fries (ESI)

Article headlines

- Potential challenges to the communication through coherence (CTC) hypothesis.
- Modulation of peak frequency in the gamma range by endogenous and exogenous factors.
- Gamma frequency reflects saliency.
- Rhythms with higher frequencies will more successfully entrain downstream populations.
- Theta-modulated gamma power as a potential mechanism of inter-areal competition.
- Top-down and bottom-up factors contribute to gamma frequency and the selective transmission of information.

Major Contents

- How is communication established?
- Does modulation of the gamma frequency preclude inter-areal coupling?
- Gamma peak frequency shifting may affect inter-areal coupling (Ray and Maunsell, Roberts, et al.).
- What effects peak gamma frequency?
- Fovea vs. periphery
- Motion vs. static
- High contrast vs. low contrast
- Attended vs. unattended
- Frequency shifts –> indicate saliency? Gamma frequency is affected by excitatory drive in local circuits (optogenetics).
- Gamma power is modulated by the phase of activity in the theta band.
- Theta band could relate to exploratory activity.

Theta modulated gamma could indicate sampling of relevant stimuli for further processing (Lisman and Jensen, Neuron 2013).

In such a scenario, theta phase periodically resets gamma – opening up channels for inter-areal communication. In this scenario, a faster gamma frequency would more robustly entrain downstream targets. I.e. the more salient item will induce a higher frequency oscillation in the lower area population, which will subsequently be preferentially transmitted (i.e. selected) by the downstream population. Winner-take-all model of feed-forward entrainment. Modelling agrees that higher frequencies win the battle for synchronisation. Top-down attentional selection may function by “priming” local population for gamma by VIP-SOM-PV inhibitory circuit. Primed circuit could maintain a higher frequency by alignment of PV cell spiking (Tsienga).

Different populations locked to different gamma rhythms.

Box 1: Essential facts

Gamma frequency is not static, but seems to be affected by extrinsic and intrinsic factors.

Dynamic frequency shifting could pose a critical challenge to inter-areal communication as hypothesise in CTC. The lack of a stable oscillation frequency could preclude inter-areal coupling.

Mounting evidence suggests that shifts in the peak gamma frequency are indicative of overall stimulus saliency.

Gamma synchronisation occurs in the absence of stimulation (vinck, lewis), i.e. by endogenous factors.

Inter-areal synchronisation appears to occur in the gamma band, despite frequency shifting.

Gama frequency may relate to overall excitatory drive in the network. Pyramidal (PING) versus Interneuron (ING) gamma in cortical circuits.

Top down signals may be realised by shifting the selected population (for location, feature, etc.) into a distinct dynamic regime that increases the local oscillatory frequency. This modulation would enhance feed-forward drive.

Cortical correlates of low-level perception: unimodal and multimodal sensory integration

Authors: Yves Frégnac, Brice Bathellier (CNRS)

What is low level perception?

Low-level perception can be defined as the computations building unconscious or self-generated inferences during the processing of sensory events. The outcome of the perceptual process is highly conditional on the context of predictive knowledge derived from the past sensorimotor experience originating from one or multiple sensory modalities. In some cases, perception departs from the physical reality (hallucinations, illusions), but this autonomous process is what ultimately feeds and guides our higher-level cortically-mediated interactions with the world and allows a stable embedding into fast but context-adaptive behavioural choices. In this article, we review the current knowledge on the structures and processes that underlie low level perception, and discuss in how far they represent crucial building blocks of any realistic model of the brain.

Describing the process of perception

When facing a natural visual scene, human subjects have an immediate conscious perception of the elementary features that compose it (segmentation), as well as of the higher-order global objects that emerge from their associations (binding), although not necessarily in this order. Contours, colours, textured surfaces, shapes, and three-dimensional objects pop out unambiguously, in a fraction of a second to seconds, according to the background context. Decomposing this evident process is a challenge for science.

Gestalt theory and theories of object recognition support two views which are not mutually exclusive: when driven by external sensations and bottom-up activation, the cortical neuronal machinery generates an automatic interpretation of our environment through built-in compositionality mechanisms, and already perceptual laws can be extracted that apply to automatised sensing; However, animals and humans cannot be reduced to the status of passive receivers facing an external physical reality; thus, other components have to be included, such as motivation, attention, expectation, action, decision, and memory. These internally generated modulations activate top-down processes, which construct unconscious hypotheses about the outer world. Consequently, the primal sketch of our sensory periphery is continuously updated and modulated by the feedback or proactive context of our thoughts, intentions, and actions. Interestingly, the cortical machinery is equipped with the appropriate substrate facilitating propagation belief across the network and turning the brain machine into a Bayesian estimator.

Subcortical and cortical circuits for sensory perception: a web of interactions

Central sensory networks appear as a major crossroads where feed-forward, and recurrent and feedback processing, merge to form a contextualised perception of peri-personal space. One of the challenges of modern neuroscience is to link the action of feed-back and crossmodal circuits to specific effects on perception.

Encoding perception: Classical receptive fields and beyond

Classical receptive fields obtained mainly with reverse correlations techniques capture in how far various attributes of the stimulus can be linearly decoded from the spatio-temporal patterns of the response. However, this LN black box approach has severe limitations and multi-scale studies of neural activity show a number of studies show that many non-linearities are embedded in neuronal spiking and subthreshold responses, which may play a crucial role for the generation of sensory representations. In addition, cross-

modal effects modulate uni-modal response, suggesting that low-level perception emerge from a complex integration of elementary sensory features.

Learning to perceive: cortical correlates

Even low level perceptual unimodal and multimodal processing are dynamically shaped by the environment. Change in the perceptual landscape produce adaptations of the cortical circuits, in particular before the critical period. This dynamical tuning of the brain to the environment might be a critical feature differentiating biological neural systems from machine.

Where models of early sensory processing succeed and where they fail.

1) Current models easily explain linear receptive fields but increasing difficulties arise to explain non-linear effects.
2) Multisensory integration is so far modelled as a read-out problem in which information from different sensory modalities is added up for optimal decisions. This fails to explain the presence of feedback and cross-modal connections, which propagate multisensory information at an early stage of processing

Conclusion:

The current observations indicate that low level perception originates from non-linear circuits with multiple nested loops. Data pinpointing the specific roles of these loops as well as models provided an integrated view of their functions are the next step for understanding the mechanisms of sensory perception in mammalian brains.

Boxes:

What we know.

Perception is an active process that emerges at the crossroad between bottom-up representations of the environment and top-down internal inferences.

Early sensory cortex circuits receive and process both bottom-up sensory information and top-down and crossmodal signals.

Neural representations of the environment in early sensory cortex already contains complex features and non-linearities, which probably reflect low level perceptual processes.

What we need to know

What is the exact impact of local recurrent and intrinsic lateral connectivity in early sensory cortical areas in the self-organisation of Gestalt laws?

What are the mechanisms responsible for the non-linear features of low-level sensory cortical representations?

What is the exact impact of feedback and crossmodal connections on cortically-mediated processes of object recognition ?

In how far can neural-based models of the sensory systems including non-linear cross-modal interactions between early primary unimodal primary sensory cortical areas explain the performances of human uni- and multi-modal perception?

Neuroscientific constraints for computational approaches of self-consciousness

Authors: Olaf Blanke (EPFL & University of Geneva), Andrea Serino (EPFL), Mel Slater (UB, UCL)

In this opinion paper Blanke, Serino, and Slater will describe the behavioural laws, the brain regions, the neuronal populations, and their interactions that subtend bodily self-consciousness and that are essential neuroscientific top-down constraints to create theoretical models and simulations of BSC.

After briefly defining consciousness, self-consciousness, and bodily self-consciousness (BSC), we will introduce the three main components of BSC that we will review: body ownership, self-location, and first-person perspective.

In the next section we will describe selected major laws of multisensory stimulus integration, the main sensory modalities and their integration principles as they are of relevance for BSC (proprioceptive, tactile, visual, vestibular), and finally in a shorter section the role of interoceptive multisensory signals in BSC. We will develop here a concept that distinguishes two important bodily reference frames (trunk-related stimuli versus hand-related stimuli), their respective spatial coordinates, and processing hierarchy, and how they link to the main components of BSC: body ownership, self-location, and first-person perspective.

This will be followed by the detailed anatomical description of human brain regions involved in BSC, and their homologous regions in non-human primates. The following regions will be reviewed, posterior parietal operculum (SII), posterior superior temporal gyrus, insula (PIVC), posterior parietal cortex (area VIP), premotor cortex as well as extrastriate body area and medial prefrontal cortex. Functional and anatomical connections between these areas will be discussed in humans (i.e. fc fMRI, DTI), and in non-human primates (i.e. tracer studies). Special reference will be given to the introduced bodily reference frames and the three key elements of BSC.

An essential part of the opinion paper will be the next section that will detail the major properties of neurophysiological responses of bimodal and trimodal single neurons in selected brain regions. We will describe these properties for three exemplary brain regions in parietal and temporo-parietal brain regions as studied in non-human primates (VIP, PIVC, SII). Special reference will be given to multisensory integration of bodily signals (proprioceptive, tactile, visual, vestibular) by these neuronal population, the two introduced bodily reference frames, the receptive field properties of these neuronal assemblies, and their processing hierarchy.

Given the current lack of data on developmental data in humans and non-human primates and the importance of novel technology developments in the study of BSC, we will describe recent methodological achievements in the fields of virtual reality and robotics. These technologies are necessary to apply highly controlled stimulus presentations in order to induced altered states of BSC during behavioural and brain imaging studies. This section will then discuss existing theoretical models and their limits based on Bayesian inference of multisensory integration in general and with respect to BSC.

We will conclude with missing data, required technological developments, lacking important experiments, highlighting the importance of computational approaches. We will close by discussing links of BSC with other cognitive architectures, covered in the special issue (such as multisensory integration, spatial navigation, visual consciousness, and the social brain).

A multi-level organisation for the sense of confidence in the brain

Authors: Florent Meyniel (CEA), Zachary Mainen (FCHAMP), Mariano Sigman (CEA/Buenos Aires)

Introduction

What do we evaluate when we assess our own confidence in a perceptual discrimination? Our ability to provide an estimate of whether our decision is correct, or our ability to estimate the level of noise in our perceptual system, or in our decision system, or in our motor-response system? Different sources of noise and different systems are intermingled in the estimation of confidence. In addition, confidence may not always be as explicit as a verbal report, it may actually be pervasively found implicitly in many behavioural and neural signals, across a variety of cognitive domains (perception, performance monitoring, learning, probabilistic reasoning, …). Therefore, in this review, confidence is recast as a general problem, not tied to a particular cognitive domains or to a particular level of processing. Based on behavioural and neural data from humans and other animals, a general notion of confidence is dissected to investigate how confidence may build up across several levels and types of information processing.

Theoretical foundations of confidence

Theories can be powerful tools for conceptualisation even if they are disconnected from the actual biological implementation (Chater & Tenenbaum 2006)[81]. The theory is considered first, the actual picture is provided in the 2nd & 3rd sections. In this first section, the aim is to concisely and quantitatively define confidence within a unified framework. The definition is general to foster a broad investigation of the notion in biological systems. In addition, the theoretical motivations are listed: in principle, why confidence could be useful for?

A unified, general framework for confidence

Distributions of probability: confidence defined as a 'degree of belief'

Probabilities are, by essence, tailored to formalise confidence. Notion of variance.

A particular but pervasive case: hypothesis testing

The simplest case: when the probability distribution comprises only two elements (e.g. correct vs. incorrect). Refer to classical statistics (Kepecs & Mainen 2012)[82].

The generic aspect of this definition

Discuss internal-external sources of uncertainty.

Give a quick tour in the computational models of confidence (reviewed in Kepecs & Mainen 2012)[82] to show how they quantify, in the end, a probability distribution.

In practice: what confidence could be useful for

Make the best choice

Uncertainty is a determinant of choice (Pascal, the prospect theory).

Optimally weighting of sources of evidence, e.g. from different sensory channels (e.g. Körding 2004 visiomotor; Ernst & Banks 2002)[83,84]; use uncertainty to arbitrate between different strategies (Daw, Niv & Dayan 2005)[85]; complement likelihood with priors (O'Reilly et al. 2013)[86].

Decide based on our errors or error likelihood

Be able to detect our errors (error monitoring literature);

Estimate our error likelihood, for post-decision wagering, for opt-out (Kiani & Shadlen 2009)[47], to choose the easiest decision (Barthelmé & Mammasian 2010)[48].

Improve learning & decision processes

Adapt to changes & reversals: identify nested levels of uncertainty (Behrens et al. 2007; Nassar & Gold 2010)[53,54]; 'risk prediction error' (Preuschoff & Bossaerts 2007)[87]; precision of cause (Vossel et al. 2014)[88].

Decision to get more information (Vul et al. 2014)[89]; foraging (Kolling et al. 2012)[90]

Improve collective decisions by sharing confidence (Bahrami et al. 2010)[91]

Characterisation of confidence in humans and other animals

The aim is to show that confidence is not just a fancy concept but that humans & other animals do compute confidence estimates. Here, we track confidence at several levels of processing, ranging from the most obvious (explicit report) to the most indirect (some signals do not reflect confidence but reveal that uncertainty is taken into account at some point of the process). We then show that confidence is not a mere, unified readout of uncertainty but instead it reflects the sometimes distorted processing of uncertainty through parallel or hierarchically organised levels of information processing.

Evidence that confidence is processed

Direct evidence: explicit readout of confidence

As a property of the stimuli & task: reading out the objective uncertainty in sensory systems (e.g. visual: Barthelme & Mamassian 2009)[46]; building an abstract, task-independent and transferable estimation of confidence (de Gardelle & Mamassian 2014)[92]

As a property of the decision & the decision maker: for subjective decisions (e.g. value-based decision: de Martino 2013)[93]; for the estimation of performance

Moderately indirect evidence: implicit metacognitive signatures

As a property of the objective difficulty or subjective performance: Opt-out task (Kiani & Shadlen 2009; Paul et al. 2011)[47,94] , waiting time (Kepecs et al. 2008)[45].

Indirect evidence: signals reflecting a confidence-based computation

Categorisation under uncertainty (Summerfield et al. 2011)[55]; Probabilistic reasoning: in infants (Denison & Xu 2014, Téglás et al. 2011) [95,96]; probabilistic reasoning & all-or-none responses (Vul et al. 2014)[89]

Brain signals modulated by uncertainty: value signals[97], pupil & reliability (Nassar et al. 2012)[98]; outcome probability[99] (Kepecs et al. 2008; van Duuren Pennartz 2008)[45,99]; outcome entropy (Monosov & Hikosaka 2013, Christopoulos et al. 2009, Strange et al.2005)[100–102]

Confidence is not a simple, unified, unbiased readout of uncertainty

Dissociating confidence from its substrate

Readout of the wrong variable (e.g. introspection vs. reasoning by Kahneman & Tversky 1982)[103]

Confidence is not just a hierarchical readout: performance without confidence (Komura 2013 pulvietnar[104]; Rahnev et al. 2012 septum[105]; blindsight); confidence without performance (Charles et al. 2013)[106]

Distinct forms of uncertainty encoded in the brain

Expected vs. unexpected uncertainty (Bach & Dolan 2012)[107], nested levels of uncertainty (Payzan-LeNestour et al. 2013)[108], learned vs. given uncertainty (Fitzgerald et al. 2010)[109]; state vs. outcome prediction errors (Gläscher et al. 2010)[110];

Noisy & biased confidence estimates

Noisy readout: Noisy representation of posterior probabilities (Acerbi et al. 2014)[111]; noisy readout of the decision signal (de Martino et al. 2013)[93]

Not just noise, but also biases: distorted perception of probabilities (Kahneman & Tversky 1979; Lench et al. 2014)[112,113]; inferential errors, e.g. law of small number (Kahneman & Tversky)[114]; Naive intuitive statistician (Juslin 2007)[52]; miscalibration, e.g. overconfidence (Mamassian 2008)[115], under-confidence

Implementation of confidence in the brain

Aim: review how the estimation of confidence may be implemented in the brain, at the mechanistic level, not by pointing out correlates of confidence in brain signals. The description should reflect the multiple levels and the multiple routes for confidence estimates in the brain.

Decision models at the neuronal level

Accumulation of evidence by neurons

The family of accumulation models, mostly for 2-alternative choices (Yeung & Summerfield 2012 review, Pleskac & Busemeyer 2010; Kiani & Shadlen 2009; Fetsch et al. 2014)[47,116–118]

Elaboration: models with multiple alternatives (Churchland & Ditterich 2012; Krajbich & Rangel 2011)[119,120]

Sampling & simulation models

Motivated theoretically by the probabilistic mental models (Gigerenzer et al. 1991, Koriat 2012)[50,121] and more recently by the sampling-based neural codes (Hoyer & Hyvärinen 2003; Fiser et al. 2010)[122,123]. Could be extended to internal sampling for value-based decisions (Krajbich et al. 2010)[124], and mental imagery for inter-temporal choices (Peters & Büchel 2011)[125]

Neural codes for confidence

Probabilistic neural code

The 'Bayesian brain hypothesis' (Knill & Pouget 2004)[126]; Neural assemblies form probabilistic population codes (Pouget et al. 2013)[127]. Optimal readout can be implemented realistically by neurons (Deneve et al. 1999)[128].

The probabilistic view does not guarantee optimality, thus, it is not disproved by the existence of biases: probabilistic reasoning + noise can lead to biases (Costello & Watts 2014)[129].

Passing confidence from one level to another: neural 'Readout'

Decision confidence readout by lateral OFC from the decision signal in the VMPFC (de Martino et al. 2013)[93].

Confidence transmitted to the outcome expectation in the OFC (Kepecs & Mainen 2009, van Duuren 2008)[99,130].

Confidence between cortical layers: precision-weighted prediction error (work by Friston).

Sharpening of tuning curve and 'gain control' (Kok et al. 2012)[131].

Neuromodulatory control of the level of noise in the network: acetylcholine, noradrenaline (Yu & Dayan 2005)[132].

Summary: key facts

- Confidence is a 'degree of belief', a notion that exists in a variety of cognitive domains
- Defined narrowly, confidence is accessible experimentally through explicit reports; more broadly, it is also reflected in implicit measures (choices, reaction times, …)
- Implicit measures extend the study of confidence to non-human animals and reveal that confidence is not uniquely human
- The degree of ratiocination in confidence estimates varies across tasks and individuals
- The estimation of confidence can be tightly related to the decision process
- Estimation of confidence can also be more indirect, relying on heuristics
- The accuracies of the calibration (the mean value) and the resolution (the trial-to-trial variations) of confidence estimate are partly uncorrelated
- Although they are often correlated, confidence can be dissociated from performance
- Confidence does not reflect a single kind of uncertainty; distinct neural representations underpin distinct kinds of confidence
- Confidence could be represented at different levels of processing: the representation of sensory information by neural codes, the decision processes, a post-decision evaluation

Consolidation: Shaping and Reshaping Memory

Authors: Jan Born (EKUT), Yadin Dudai (WIS), Avi Karni (UHAIFA), Rony Paz (WIS)

Introduction

Definition and taxonomy of memory consolidation (immediate/cellular, delayed/systems). Consolidation following one-shot vs. repetitive and declarative vs. non-declarative encoding. Time-line and critical 'decision' nodes ('rites of passage'): before encoding (schemata, anticipation, context and valence); in and immediately after encoding; throughout the consolidation window; retroactive interference; the role of sleep; 'reconsolidation' in reactivation. The importance of understanding consolidation in modelling realistic memory systems and in designing high-fidelity cognitive learning systems.

Multi-level analysis of consolidation

- Postulated computational goal(s) of memory consolidation (e.g. pruning, integration, updating, parsimony).
- Algorithms used in implementing the postulated goals:
  - in immediate ('synaptic') consolidation (e.g. rendering the system immune to the consequences of molecular turn-over)
  - in delayed ('systems') consolidation (e.g. complementary learning systems, CLS, to permit parallel encoding with consolidation, pruning and integration; offline reactivation in sleep)
- Implementation of the posited algorithms (including description of brain circuits and declarative - non-declarative interactions)

Our approach

Four brief examples of experimental approach and findings, citing published papers:

- Born (transfer of information from hippocampus to extrahippocampal cortex in sleep-dependent consolidation)
- Dudai (immediate binding and triggering of consolidation in one-shot experience)
- Karni (role of motor cortex in skill consolidation)
- Paz (effect of valence on consolidation and integration of new into existing information).

Conclusion: Missing data, open problems ripe for experiments and modelling.

Box “what we know”: Synaptic and cell-wide underpinning, role of MTL, role of PFC, CLS, role of sleep in multiple memory systems, reconsolidation

Box “what we need to know”: Brain mechanisms by which existing knowledge, anticipation and ambient state, valence and context shape consolidation. Role and mechanisms of implicit reactivation in reconsolidating and reshaping memory. Boundary conditions on consolidation and reconsolidation of one-shot and repetitive experience. Does consolidation ever ends? Are there ways of enhancing consolidation, and will this increase memory capacity? Effect of aging on consolidation.

The cognitive architecture of working memory

Authors: Johan Eriksson (UMU), Ed Vogel, Fedrik Bergström (UMU), Lars Nyberg (UMU)

Introduction

Multiple ways of defining WM, we here focus on WM as [to be agreed upon]; purpose/aims of the paper (describe cognitive architecture of WM, defined as "a description of the brain regions and their interactions that subtend a specific cognitive function in humans or in behaving animals"); relation between attention and WM, WM as activated LTM

Major behavioural laws

Capacity limits (including individual differences, "continuous vs. discrete resource allocation", chunking); distraction/interference (necessity of refreshing); modularity/representational domains (verbal, spatial, objects)

Brain areas and connections

Combining neuroimaging and lesion/patient/TMS studies (animal studies?), organising regions according to core WM processes (e.g., encoding, maintenance, manipulation, retrieval, updating); DTI/tracing studies

Neuronal codes, major properties of neurophysiological responses

Sustained activity during delay; Mongillo et al.'s "synaptic" theory[133]; "mixed selectivity"/high-dimensional representations (Rigotti et al. 2013)[134]; neurotransmitters (D1, D2, including NMDA receptors (?)); EEG/LFP oscillations

Plasticity, development, and learning

Lifespan changes, WM training

Theoretical models and their limitations

There are many models… We should consider limiting this in some way. One aim with the paper is to point out important model elements to consider in future attempts to formulate WM models, especially in models attempting to capture all key aspects of WM rather than specific components/parts.

Conclusion: missing data, required experiments, open problems ripe for modelling, etc

Box "what we know" (see below – "Neurocognitive constraints")

Box "what we need to know": The relation between WM and consciousness, the nature of WM representations/representational "bottlenecks", genetics, "integrating" models

Neurocognitive constraints for working memory (draft)

1) Accessibility – The representations held in working memory are accessible for use in relation to other processes, such as planning, decision-making, etc. Working memory can be conceptualised as providing an interface between perception, long-term memory, and action.
2) Durability – The representations held in working memory last for several seconds to minutes without external support. Durability relies on active processes such as rehearsal ("refreshing" the content), without which the representations decay within a few seconds.

3) Capacity constraints – Working-memory capacity is limited and may only contain 1-4 items (absolute limits remain controversial). Capacity can be increased through "chunking" bits of information into more complex units.
4) Modularity – Working memory is not a unitary process, but results from the interaction between several relatively separable modules. Key modules (highly schematic) are a) item representation modules that temporarily stores working-memory content, and b) attention/executive modules that control the content of working memory through processes such as selection and updating, interference resolution, etc.
5) Flexible representations – The content of working memory can be manipulated, such that the content itself can be "worked on" and changed, e.g., during mental arithmetic.
6) Distraction/interference – Working memory processes are sensitive to interference, which is a major factor for working-memory failure.
7) Input – The content of working memory can come from several different sources, e.g., visual or auditory sensory, but also from long-term memory.
8) Brain regions – Several different brain regions contribute to working memory, reflecting its modular nature.
    a) Prefrontal cortex – Persistent neural activity during working memory has been demonstrated consistently and is likely to reflect goal-related, cognitive control representations, although traditional interpretations suggest that PFC activity reflects working-memory content.
    b) Parietal cortex – Attention and spatial representations.
    c) Sensory cortex – Recent research suggest that working-memory content is represented by activity in sensory regions; the specific region depends on the specific content.
    d) Basal ganglia – Suggested important for updating working-memory content.
    e) Cerebellum – Verbal working memory
9) Neurotransmitters – Dopamine etc.
10) Working memory does not seem to require structural alterations such as new protein synthesis, as it works with pre-existing representations ("activated LTM")

The cognitive architecture of spatial navigation: hippocampal and striatal contributions

Authors: Fabian Chersi (UCL), Neil Burgess (UCL)

Introduction

Why spatial navigation is a good model for understanding general issues in cognitive neuroscience. This is due to knowledge of the neural representations and learning rules involved, their distribution in the brain and correlation of navigational behaviour with local damage, inactivation, or measures of neuronal activity. Here we examine the cognitive architecture of spatial navigation, with a focus on hippocampal and striatal systems, and their interaction. We outline how this cognitive architecture could be implemented in simulations of the mechanisms driving navigational behaviour at the level of neuronal firing rates.

Multiple representations support spatial navigation

Outline the types of representation that would be useful for spatial navigation from a theoretical perspective (allocentric, egocentric, trajectory-related, reward-related).

Briefly review behavioural evidence for multiple representations of location that are active in parallel and combine to support behaviour, see e.g. the behavioural studies reviewed in (Burgess, 2006)[135].

Briefly review the various neural representations relating to spatial navigation found in freely moving rodents: head direction cells (throughout Papez's circuit, including medial entorhinal cortex mEC and the subicular complex SC), place cells (hippocampus), grid cells (mEC and SC), boundary cells (mEC and SC), trajectory cells (parietal cortex), trajectory/reward-approach representations (striatum). Recent reports of object-location trace cells in lateral EC.

Outline current thinking on how these spatial representations interact (see e.g. introductory review by Barry & Burgess, 2014)[136]: head-direction cells set the orientational framework for place, grid and boundary cells, driven by distal cues, if present; boundary cells determine relation of place cell firing patterns to the environment, place and/or boundary cells determine relation of grid firing patterns to environment; trajectory cells and head direction cells are required for grid cell firing; grid cell firing may aid place cell firing by providing a path integration input.

Systems neuroscience of spatial navigation

Briefly review experiments in which local lesions or inactivation produce specific changes in navigational behaviour in rodents, specifically the plus maze task described by Packard and McGaugh (1996) [137] the variation on the Morris watermaze task described by Pearce et al. (1998) [138].

In the Packard and McGaugh experiment, rats were trained to approach a consistently baited arm in a plus-maze from the same start arm. After several days a single probe trial was given, in which rats were placed in the start arm opposite that used in training and allowed to approach the central decision point. Control rats displayed 'place learning' (i.e. going to the same place in the room) on the Day 8 probe trial and 'response learning' (i.e. making the same body turn) on the Day 16 probe trial, indicating that with extended training there is a shift in the system controlling behaviour. Supporting this interpretation, rats with inactivation of striatum displayed place learning on both Day 8 and Day 16 probe trials, whereas rats with inactivation of the dorsal hippocampus showed no preference for place or response learning on the Day 8 probe trial, but displayed response learning on the

Day 16 probe trial. Thus, it seems that response learning depends on the striatum while place learning depends on the hippocampus.

In the Pearce et al task, a submerged escape platform is placed in specific locations of the pool with a visible landmark located nearby, and rats gradually learn relatively direct paths to the goal over the course of a few trials. After four trials (one session), the escape platform and the landmark are moved together to a new location in the maze. Rats with and without hippocampal lesions both are able to reach the hidden platform but present specific and distinct performance curves. Hippocampal lesion animals quickly locate the platform on the first trial of a new session, using the intramaze landmark as a cue, whereas than the control animals are slower, continuing to search at the previous location in the maze. But the control animals learn the new location within each session, and out-perform the lesioned animals by the fourth trial of the session. Thus, the hippocampus appears to support learning of the platform location relative to the maze – using the boundary of the maze in combination with distal cues for orientation (e.g. experiments by Hamilton, Sutherland et al., 2007)[139].

Briefly review related experiments in humans using functional neuroimaging during virtual reality navigation.

Corresponding to the Packard and McGaugh (1996)[137] experiment, correlation of caudate activity with navigational performance is seen during well-practised (Hartley et al., 2003)[140] or otherwise stereotyped (Iaria, Bohbot et al., 2003)[141] route following. Equally, correlation of hippocampal activity with navigational performance is seen during unconstrained way-finding with respect to environmental cues within a previously explored environment (Maguire et al., 1998; Hartley et al., 2003; Iaria, Bohbot et al., 2003)[140–142].

Corresponding to the Pearce et al (1998) experiment[65], experiments have examined the learning of object locations within in a virtual environment containing an intramaze landmark, an enclosing boundary and distal orientation cues. In this task, hippocampal activity corresponds to learning object locations relative to environmental boundaries, whereas learning object locations relative to an intra-maze landmark corresponds to activity in dorsal striatal and parietal areas (Doeller et al., 2008)[143]. These findings are consistent with the identification of hippocampal representations with place cell firing driven by distances to environmental boundaries within an orientational framework determined by distal cues. The precise representations supporting the striatal/parietal contribution to navigation are not known, but likely involve trajectories relative to the intramaze landmark.

Learning rules and control strategies in spatial navigation.

There is a long history of debate concerning the nature of spatial learning, spanning from the proponents of stimulus-response associative learning mechanisms driven by trial and error to the proponents of incidental learning of internal representations capable of supporting cognition (see e.g., Hull 1943; Tolman, 1949)[144,145]. These arguments are brought to current thinking on spatial navigation in terms of reinforcement learning based on prediction error (Rescorla & Wagner, 1972; Sutton & Barto, 1981; Foster et al., 2000)[146–148] and the proposition of the hippocampus as a ‘cognitive map’ (O’Keefe and Nadel, 1978)[149].

Briefly review the examination of blocking and overshadowing between boundaries and landmarks in learning object locations (Doeller & Burgess, 2008)[150], and its conclusion that striatal (landmark-action) learning relies on prediction error, while hippocampal (boundary-place) learning is incidental.

Briefly review the literature on combining hippocampal and striatal systems in navigation, possibly mediated by medial prefrontal cortex (Killcross & Coutureau, 2003; Miller &

Cohen, 2001)[151,152]. Discuss the supporting evidence from Doeller et al. (2008) [143] in which the more active system controls behaviour, and increased medial prefrontal activation is seen when the balance between striatal and hippocampal activity is closest.

A neural systems account of spatial navigation

In this section we outline a computational model of how the hippocampus and the striatum might support spatial navigation in the two experiments reviewed above, the modified Morris watermaze task described by Pearce et al. (1998) [65], and the plus maze tasks described by Packard and McGaugh (1996) [137].

The proposed model uses the local, incremental, and statistically efficient temporal difference learning rule for the striatal component and a one-shot learning rule for the hippocampal formation. The first is a reinforcement-based "actor-critic" network that is a general model of classical and instrumental conditioning. In this case, it is applied to navigation, using sensory input to provide information about state. By itself, the actor-critic can efficiently learn to solve both the water and the cross maze tasks, but its learning is slow and is prone to errors when the sensory state ambiguous like in the cross maze task. In addition to this system, we argue that the hippocampus possesses a goal-independent representation of space that is learned very rapidly, perhaps relating to its role in episodic memory. This is obtained by having place cells driven by their proximity relation to environmental boundaries (see e.g. Hartley et al., 2000)[153], and then using rapid Hebbian learning to associate place cell firing to 'goal cells' whose firing will provide gradients that can be used for goal-directed navigation (see Burgess & O'Keefe, 1996; Foster et al., 2000)[62,148].

Each component of the model is necessary at a different stage of the task: initially a rapid (but imprecise) goal-directed learning method, and thereafter a stimulus-response mechanism that is capable of learning the statistics of the task. The final component of the model, potentially corresponding to medial prefrontal cortex, is involved in selection of either the hippocampal or striatal system to control behaviour (see also Chersi & Pezzulo, 2012; Dollé et al., 2010; Sheynikhovick et al., 2009)[154–156]. The mechanism for selection compares the 'confidence' of signalling goal direction by either system, which is measured in terms of the local gradient of the normalised value function expressed by each system.

Preliminary simulations indicate the model successfully captures gradual acquisition in both tasks, and show good matching with experimental data on learning rates in control and inactivation/ lesion animals in both of experiments.

Box: main facts

1) Multiple spatial representations have been identified in neuronal firing, and in behaviour.
2) There is a good mapping between representations and brain systems.
3) These systems appear to combine constructively to support spatial memory, which implies that they can be selected between in an appropriate manner, e.g. according to a measure of 'confidence' in each system (e.g. slope of value function).
4) Different systems appear to use different learning rules, potentially reflecting optimisation for different aspects of the task (1-shot learning for hippocampal episodic memory, prediction error for striatal action learning)
5) Open questions include: the representations in striatum and parietal areas that could support e.g. landmark-related and response learning, are they the same, are there multiple representations, and if so what are they like? How does the output from the hippocampal formation interact with reward-related representations in the ventral

striatum to which it projects? What is the nature of the incidental learning rule posited for the hippocampus – does novelty detection play a role similar to prediction error in the striatal system, and if so is this signalled by dopamine?

Varieties of sequence knowledge: From transition probabilities to symbolic rules and linguistic trees

Authors: Stanislas Dehaene (CEA), Florent Meyniel (CEA), Catherine Wacongne (CEA), Liping Wang (CEA) and Christophe Pallier (CEA)

Introduction

When exposed to temporal regularities in their environment, humans and many other animal species are able to internalise them in the form of internal models and to use these models to generate expectations on futures states. Such sequence knowledge is most evident in the complex nested structures of human language, but even non-human primates can encode abstract temporal structures. Here we review what is known about the neuronal representation of sequence knowledge in human and non-human primates. We argue that there is evidence for a multiplicity of parallels systems, each capable of representing knowledge of the incoming sequence at an increasing degree of abstraction. We propose, at a minimum, a distinction between five different types of internal representation: timing knowledge; ordinal knowledge; chunking; symbolic rules; and recursive tree structures. In each case, we first introduce a precise vocabulary and theoretical framework that isolates the specific type of sequence knowledge involved. We then give examples of experimental paradigms and neurophysiological data, determine the behavioural and neural signatures of the systems involved, and discuss whether any of these levels is unique to humans.

Level 1. Timing knowledge

Many animal species are able to represent time intervals and use these temporal representations in simple computations. The representation of time is a major topic of research, for which several reviews are available. Here we focus solely on their contribution of to the internal representation of temporal sequences.

Example 1: Gallistel's temporal choice task (Balci et al. 2009) [157]. Example 2: the mismatch negativity and the effect of interstimulus interval, omissions, and duration mismatches.

Proposal: what is being encoded here is the probability of a specific transition between items at a specific time (Wacongne et al. 2012) [158].

Basic signatures of the system: (1) Weber's law. (2) prediction and prediction error.

Different types of mismatch responses can be explained by replication of the same mechanism at different cortical levels. Thus, the neural mechanism is likely to be highly distributed in many sensory systems, motor and decision systems. Give examples in several domains: auditory, visual (Gavornik & Bear 2014; Meyer & Olson 2011)[159,160], motor conditioning, and reward anticipation (Fiorillo et al. 2008)[161].

Role of subcortical structures such as the basal ganglia and the cerebellum (Meck, Ivry; evidence from Parkinson and Huntington diseases). The hippocampus may also encode elapsed time (Kraus et al. 2013)[162]. A parietal network may be involved in time representation, comparison and computations (Leon & Shadlen 2003)[163].

Neurophysiological data and hypotheses on timing (Eagleman et al. 2005; Jin et al. 2009; Laje & Buonomano 2013)[164–166].

Level 2. Ordinal knowledge

While the level 1 system generates precisely timed predictions, others systems abstract away from timing information and encode only temporal order. Working memory may have evolved as a system for predicting *that* something will happen without knowing *when*.

Humans : dissociation between MMN and P3 based on the resistance to changes in inter-stimulus interval (Pegado et al. 2010)[167]. Review other evidence for a reduced or absent MMN when the timing is jittered.

In monkeys, the paired-associate paradigm suggests that long-term expectations pass through the prefrontal cortex (Asaad et al. 1998; Takeuchi et al. 2011; Tomita et al. 1999)[168–170].

Monkeys can learn to represent the ordinal parameter of sequences (Orlov et al. 2000; Terrace et al. 2003)[171,172].

Some PFC neurons encode temporal order in their firing rate (Jacob & Nieder 2008; Ninokura et al. 2004)[173,174]. Even a binary code has been observed (Shima & Tanji 2006) [175].

Some PFC neurons also use phase coding to encode memory for temporal order(Siegel et al. 2009)[176].

Level 3: Chunking and the formation of hierarchies

The frequent occurrence of reproducible series of items in a sequence allows the nervous system to "chunk" these items, i.e. create a single entry for them in an internal "lexicon".

Chunking was first introduced in working memory (Mathy & Feldman 2012; Miller 1956)[177,178].

Human infants chunk sequences of syllables into putative words (Saffran et al. 1996)[179]. Critique of most experiment: most of the data can be explained by statistical learning of transition probabilities. However, there is some evidence that children do form « words » that they can attach to objects (Graf Estes et al. 2007)[180]. This ability may even be accessible to other animals (Kaminski et al. 2004)[181].

Evidence for chunking in simpler non-linguistic paradigms: Local-global paradigm (Bekinschtein et al. 2009)[182], which can be acquired by monkeys (Uhrig et al. 2014)[183].

Neurophysiological data on chunking: role of the PFC and basal ganglia (Fuji & Graybiel 2003; Graybiel 1998)[184,185].

Level 4. Symbols and rules

Sequences may be represented at a more abstract level, as transitions, not only between specific items (level 2) or between chunks (level 3), but between *categories* of stimuli.

Example:

- The formation of symbolic rules is essential to language representation, even at a phonological level (phonotactic rules governing the formation of CV structures in language) (Jacquemot et al. 2003)[186].
- Infants can learn the abstract "algebraic" structure of sequences of syllables (Marcus et al. 1999, 2007)[187,188]. This provides evidence that they use abstract categories of "same" and "different" to encode the sequences.

Neurophysiological evidence? Very scarce, but Huntington's disease patients may show a specific impairment (Teichmann et al. 2005)[189].

This algebraic level of representation of symbols and rules is available to non-human primates (Shima et al. 2007)[190]. Nieder, Miller. Even bees may represent the abstract concept of identity (Avarguès-Weber et al. 2012; Giurfa et al. 2001)[191,192]. However, primates are generally very extensively trained. Furthermore, even after such training, there is evidence that rules are not systematically applied (Yang 2013)[193].

Level 5. Recursive tree structures

Chunking and symbolic rules are not sufficient to account for the human knowledge of language. Most linguists argue that human linguistic structures require a representation of recursive rules and embeddings, with long-distance dependencies (Chomsky). What is the signature of recursion ? The formation of chunks with the same (recursive) type as the internal items. This is a key difference between phonology and syntax.

The importance of recursion is still debated(Frank et al. 2012)[194], but this could be because some simpler systems (level 2, 3 or 4) routinise the structures of language. Example of self-space reading, where eye movements may well be determined by a simpler non-recursive representation of linguistic expectations.

Neural representation of the constituent structure of language in humans: the cardinal network involving pSTS and inferior frontal regions (Musso et al. 2003; Pallier et al. 2011)[195,196], plus temporal pole. The signatures of tree structures in the language domain also include additional properties such as linguistic movement and long-distance dependencies (Shetreet et al. 2009)[197]. The dynamics of this circuits suggests a slow process of tree accretion (Pallier et al. 2011; Vagharchakian et al. 2012)[196,198]. MEG and intracranial recordings suggest a progressive build-up of power in the beta and gamma bands (Bastiaansen et al. 2010)[199].

There is evidence that syntactic tree structures exist outside of language (in humans) and may involve a similar circuitry (Bahlmann et al. 2008)[200]. However, there is also evidence that the tree structures in mathematics, music, or decision making may rely on distinct systems (Maruyama et al. 2012)[201].

Is recursion a uniquely human ability? Discuss Povinelli and Hauser's idea that this is unique to humans (Hauser et al. 2002; Penn et al. 2008)[202,203]. There is evidence for decision trees in animals, but this is hierarchical but not necessarily recursive. There is some behavioural animal data attempting to look at language-like grammars A^nB^n (Abe & Watanabe 2011)[204]. Discuss the problems with these experiments (Beckers et al. 2012)[205].

Table. The five types of sequential knowledge: definition, characteristic signatures, and associated neural systems

Box. Mathematical frameworks for sequence knowledge

We will briefly describe in this box several frameworks for modelling the acquisition and representation of sequence knowledge: (1) the predictive coding framework; (2) the mathematical theory of Kolmogorov-Chaitin complexity ; (3) the classification of grammars according to the Chomsky hierarchy, and its probabilistic (Markov) and Bayesian extensions (Tenenbaum, Goodman).

Box. What we know

Timing information is represented in multiple circuits including the prefrontal cortex, basal ganglia, cerebellum and hippocampus.

The same computational principles of prediction and prediction error underlie sequential knowledge in many different domains and modalities, and are replicated in multiple cortical and subcortical networks.

Monkeys and humans share a capacity for representing the abstract structure of temporal sequences.

Timing, ordinal knowledge, chunking, and symbolic rules are available to human infants during the first months of life.

A reproducible network (figure xx) underlies the representation and manipulation of linguistic tree structures in humans. This network is already activated during speech listening in 2-month-old infants.

Box. What we need to know: open questions for future research

- How is predictive coding implemented by neural tissue? What is the specific contribution of various cortical layers, and whether the same predictive structure is replicated in different cortical areas with different codes.
- The respective role of cortical and subcortical structures in sequence coding (cerebellum, basal ganglia).
- The neurophysiological coding of nested or recursive structures
    - A specific proposal : tensor products, and the need to test it
- What makes human representations of symbols and rules unique
    - Particularly developmental investigations of the language system
    - Is there a domain-specific system for language processing? Or, does the human uniqueness from a broader, possibly domain-general device for representing abstract rules?

Tentative list of figures

Figure x. A simple auditory test demonstrates that the brain encodes sequences at multiple levels. Example of the local-global test [206]: models and corresponding data (showing three hierarchical levels : sensory adaptation, transition probabilities, and higher-level rules)

Figure x. Evidence that non-human primates can represent abstract ordinal information [171,173,174].

Figure x. Chunking of sequences in prefrontal cortex and basal ganglia [184].

Figure x. Neurophysiological evidence that monkey prefrontal neurons encode the abstract symbolic structure of temporal sequences [190].

Figure x. Key regions of the human brain involved in the constituent structure of language [196].

Figure x. Cardinal circuit for syntactic representation and manipulation in humans: posterior superior temporal sulcus and inferior frontal gyrus (composite image from multiple converging publications).

Centrality of social interaction in human brain function

Authors: Riitta Hari, Linda Henriksson and Lauri Parkkonen

Summary

People are embedded in social interaction that shapes their brains during the entire lifetime. Instead of building on lower-level cognitive functions, social interaction could be the default mode via which the organism communicates with the environment. If this hypothesis turns out to be true, it would have drastic theoretical implications on how we think the brain works and how we should "build" or simulate the human brain (as in the HBP). Time-resolved simultaneous recordings from the brains of two interacting subjects are needed to experimentally solve this question. Here, we discuss the background, realisation and implications of such dual brain recordings.

Introduction

Social interaction is among the most complex functions humans (and their brains) perform. Yet, the interaction appears surprisingly easy. For example, during conversation, turns are usually taken effortlessly, smoothly and in a temporally accurate manner so that, over different languages and cultures, the gaps between the turns are typically only a few hundred milliseconds, or even less (Stivers et al., 2009)[207]. Such brief intervals cannot reflect just reactions to the end of the previous speaker's utterance; instead the conversation participants have to predict when the previous speaker is going to finish her turn of talk.

The importance of social interaction becomes evident from the multitude of social accounts in our everyday life. For example, we teach, learn, communicate, treat, and deceive. We punish others with social isolation. We often feel to be in the shoes of the others, feeling sad or happy with them. Many everyday social interactions are indirect, for example, when we think about the motivations of a person who just sent us an angry email.

It is also clear that "person stimuli", such as faces or bodies, are not only visually complex patterns of features but our interpretations of them go far beyond the immediate information given. Moreover people mutually adapt to the rhythm of other's movements (such as walking or speech), and behavioural synchrony is the key in various group performances in music and sports, for example.

Interaction between people takes place all the time and it is expected. For example a young baby gets nervous very soon after the mother "freezes" her face. In addition, children with Moebius syndrome, suffering from congenital bilateral facial nerve paralysis, experience difficulties in social interaction as their caregivers cannot respond properly to the child's emotions as e.g. a smile is hidden on such a face.

It has been suggested that the complexity of social relationships—that can only take place or arise in a group of people—may have driven the evolution of the human brain. Importantly we are here dealing with a group property rather than just an individual personality trait.

Many philosophers have emphasised the importance of other people to the development of self (e.g. Nietsche: "Das Du ist älter als das Ich."). Although we do not (fortunately) have any data of people who would have developed without any human interaction, impoverished interaction in e.g. orphanages has led to autistic-type serious behavioural disabilities.

Trevarthen (2010) has insightfully stated that "Infant human beings imitate other humans, not just to act like them, but to enter into a communicative and cooperative relationship with them by some transfer of the feeling of body action".

Spectator science?

Recent flourishing of social neuroscience has importantly enlarged and deepened our understanding about social stimuli, tasks, and context that affect human brain function. Especially revolutionary has been the research on mirroring mechanisms, starting with the discovery of mirror neurons in the monkey frontal cortex (di Pellegrino et al., 1992; Rizzolatti and Craighero, 2004)[208,209]. These studies have brought together various disciplines interested in the functions of human mind.

Despite of all this enormous progress, it has to be emphasised that most of the studies of social cognition and interaction have so far been based on the assumption that humans (and their brains) are reactive, and thus the ongoing state-changes triggered by the interacting Partner, or the physical environment, are not taken into account. Instead, the subject is exposed to a range of crafted stimuli from checkerboard patterns to movies.

However, this kind of research can be considered to represent *"spectator science"* where the state of the brain under study is considered to be affected by the external stimuli, only, whereas in reality the internal state of the brain changes all the time. People are participating in the events, not only observing them. Time-resolved simultaneous measurements of two persons could remediate this caveat.

Private brains, shared world

We humans have private brains but we share our world, more so with people close to us. It is thus not a trivial question to ask how we, with our own brains that are known to differ extensively in their details, can understand each other. Such mutual understanding requires a between-participant similarity in various mechanisms of perception and action (Hari et al., 2013)[210]. Accordingly, activity of some brain areas seems to "tune in" during similar sensory-affective stimulation e.g., while watching movies (Hasson et al., 2004; Malinen and Hari, 2011)[211,212] or during conversation (Stephens et al., 2010)[213].

One-person versus two-person neuroscience (1PN vs. 2PN)?

We have proposed "two-person neuroscience" as an approach to study the physiological basis of social interaction (Hari and Kujala, 2009)[214]. The experimental goal of 2PN is to differentiate interactive vs. reactive states and to understand, e.g., the easiness of social interaction. We consider 2PN setups necessary for studies of real social interaction where people mutually regulate their dynamic coupling, co-adapting their behaviours. Whether we should take the 2PN approach instead of 1PN approach depends very much on the timing of the behaviour we are studying, and phenomena where information is exchanged between the participants in time intervals shorter than 100 ms would need time-resolved brain-imaging methods, such as MEG or EEG. Topics to be studied:

- *Synchrony* of motor actions, including the inhibition of excessive imitation (Hari et al. 2014)[215], and the synchrony of eye gaze and gestures. These can be studied both in 1PN and 2PN settings, depending on the question and the timing.
- *Effect of presence.* We behave in a totally different manner when being alone and when knowing that someone else is in the same space able to perceive us. Can be studied in 1PN settings but 2PN will give access to the effects of mutual adaptation.
- *Emotional contagion*. Can again be studied in 1PN settings but 2PN allows to unravel mutual regulation of emotions—a key feature in e.g. baby–caregiver interaction.

- *Social error monitoring*. A truly social phenomenon. Representation of others' action and errors in e.g. medial prefrontal cortex; cf. monkey electrophysiology (Yoshida et al., 2011)[216].

Two-person "hyperscanning"

Motivation. Natural social interaction comprises unique events whose exact content and timing is usually unpredictable. Therefore, the same interaction sequence cannot be repeated in order to record the brain activity of the other interactor. If we only have data from one interactor, external features of the interaction would need to be quantified to be able to isolate the contributions of the brain processes supporting the interaction. However, if we have data from both brains, we can seek for correlations between the two sets of brain signals, without explicit reference to the external events.

Experimental setups.

[This section will include detailed descriptions of the 2PN-setups, and 2PN-experiments]

Our *dual-coil for fMRI scanning* enables simultaneous measurements of brain activity from two individuals (Malinen & Renvall, OHBM 2012). Benefits of fMRI in the study of social interaction. Spatial localisation. Subjects in the same scanner ~ actual presence of the other person.

We have realised the simultaneous *MEG-to-MEG recordings* between MEG labs 5 km apart in Finland, and between one lab in Finland and the other in Belgium. Benefits of MEG? High temporal resolution. (Baess et al., 2012)[217]

Data analysis of 2-person recordings

Analysis of the 2PN-data, especially obtained by means of electrophysiological methods, is still extremely challenging. Possible approaches:

1) Hyperconnectivity: functional connectivity measures not within one brain but between the two measured brains. Temporal correlations and adaptive time lags between homologous regions in the two brains, with time-dependent changes. Intersubject correlations and coherence.
2) Statistical methods addressing signals from both brains simultaneously, e.g. ICA on joint data: searching for functional networks from the two instead of one brain. The analysis is expected to pool e.g. the listening network in one brain with the speech production network in another brain. But this seems not to be so straightforward
3) Correlations with external/peripheral measures. Example corticovocal coherence between the fundamental frequency of the speaker (picked up with an accelerometer from the throat) vs brain activity of the listener.
4) Connectivity between (1) homological brain areas between brain1 and brain2, (2) connections from one area in brain1 to any area in brain2, (3) connectivity of one area in brain1 to a network of areas in brain2, and (4) connectivity between networks in the two brains. The methodological challenges naturally increase from (1) to (4). Extra difficulty is introduced by the analysis of time-varying modulations of the connectivity.

Social interaction—the brains default mode?

Philosopher R.T. Allen (1990) stated that *"We first take the world to be animate and expressive and then learn to de-personify and de-animate parts of it"*. This default working mode goes very well with the early findings of Heider and Simmel (1944) that people easily attribute mental states to inanimate objects, such as moving geometrical shapes. Moreover, people seem to have an innate tendency to see agency in inanimated objects. For example, supernatural believers see "signs" in emotionally charged non-living

objects/scenes whereas sceptics actively inhibit this tendency (reverse inference from fMRI studies; (Lindeman et al., 2013)[218]

In a similar manner, we may have an inborn tendency to interact and synchronise with others as the default mode. The origin of this tendency is easy to understand in mammals who are entrained with the mother's motor and vocalisation rhythm already in the womb. Synchrony of other persons' bodies continues throughout the adult life.

Combining motion capture data to brain imaging might allow to combine brain and body that both tend act in synchrony with the other person during social interaction.

Future directions of 2PN?

Tentative table. Benefits of 2PN in different research questions

Box. What we know

1) Regularities of the external world are printed into our brains. Other people, with their characteristic spatiotemporal appearance, are part of our environment.
2) Nervous systems developed for movement and prediction (with a special sensitivity to changes and social stimuli)
3) Bodies are involved in cognition and emotions (Nummenmaa et al., 2014a)[219]
4) Timing determines when 2PN brain imaging cannot be replaced by 1PN settings
5) Effect of emotions on common ticking of single brains, e.g. during movie viewing (Nummenmaa et al., 2012; Nummenmaa et al., 2014b)[220,221] or perspective taking (Lahnakoski et al., 2014)[222]
6) Social environment is decisive for human development; can leave even structural traces
7) Excessive imitation can be suppressed during social interaction (Hari et al., 2014)[215].

Box. What we need to know: open questions for future research

1) Interactive vs reactive mode of brain function.
2) Novel analysis methods to capture the interactive mode
3) Different timings of the theory-of-mind and mirroring systems to find out whether they together represent a dual-process mechanism of social cognition and interaction.
4) Different time spans and integration windows of social cognition
5) How to take into account the constraints posed by both timing and social interaction on the build-up and simulation of the human brain (like the simulations in the HBP)

Tentative list of figures

Figure x. 2PN MEG2MEG and dual-coil fMRI

Figure x. xxxx

Annex D: Strategic experimental protocols

Mapping individually-specific neuro-cognitive biases in the human brain

Malach group

Goals

Our goal is driven by our recent hypothesis[1] (Harmelech and Malach, TICS, attached) termed: spontaneous trait reactivation, suggesting that the patterns of spontaneously emerging coherent patterns contain information that goes beyond the mere gross-anatomical connectivity. In fact, previous results in our labs and others are increasingly supportive of the notion that such patterns can be decoded to reveal the strength of synaptic (not necessarily direct) connections between cortical sites in the human brain. It is important to emphasise, first, that such synaptic biases are likely to underlie a central source of intrinsically stored information that allows cortical circuits to decode and predict so well in the extremely noisy and rapidly changing environment. Second, that such individually, and network, specific information complements the type of statistical information that can be derived from the current simulation approaches proposed in HBP.

Specific goals will be to examine spontaneous patterns under "resting state" condition and a simple control task (tone detection). Our preliminary results demonstrate that such auditory task allows mapping of spontaneous patterns as well as resting state. Critically these patterns will be compared to coherent patterns that emerge under naturalistic sensory stimulation -simulated in our localiser by an engaging movie segment. We expect this localiser to reveal rich and new information about individual patterns of synaptic connectivity biases across the entire sensory cortex - as they are revealed during naturalistic sensory processing on the one hand, and when they are reflected in spontaneously emerging patterns.

Stimuli and Task

The localiser consists of six fMRI scans lasting a total of 56 minutes including:

- Two resting state scans (RS, subjects rests in the magnet with eyes closed) of 8 minutes each.
- Two "tone detection" (TD) scans of 8 minutes each- a brief tone is presented at random times- with 3-9 seconds inter-stimulus interval. Subject's task is to press a button each time they hear the tone.
- Two repeats of an engaging movie segment of 12 minutes each (MV). The task of the subject is simply to attend to the movie and remember its contents. Following the movie presentations subjects will undergo a memory questionnaire about the movie contents.

The localiser will start with either a RS or TD experiment (counterbalanced across participants) followed by one MV followed by two more RS / TD experiments (counter balanced) the second (identical) MV and finally the last RS/TD experiment.

Results

Our preliminary results in running this localiser reveal a remarkable level of detailed information about connectional biases. In particular we find that the spontaneous patterns do not depend on the commonly used and often criticised "resting state" and can be obtained using a highly controlled, albeit simple detection task. Second, they reveal both striking similarities and consistent differences in the patterns evoked by naturalistic

stimuli and those emerging spontaneously. We think this localiser could provide important information that, critically, is not addressed by the detailed statistical simulation proposed in other pillars of the HBP.

TICS-1250; No. of Pages 10 ARTICLE IN PRESS

Opinion

Special Issue: The Connectome

Neurocognitive biases and the patterns of spontaneous correlations in the human cortex

Tal Harmelech and Rafael Malach

Neurobiology Department, Weizmann Institute of Science, Rehovot, Israel

When the brain is 'at rest', spatiotemporal activity patterns emerge spontaneously, that is, in the absence of an overt task. However, what these patterns reveal about cortical function remains elusive. In this article, we put forward the hypothesis that the correlation patterns among these spontaneous fluctuations (SPs) reflect the profile of individual *a priori* cognitive biases, coded as synaptic efficacies in cortical networks. Thus, SPs offer a new means for mapping personal traits in both neurotypical and atypical cases. Three sets of observations and related empirical evidence provide support for this hypothesis. First, SPs correspond to activation patterns that occur during typical task performance. Second, individual differences in SPs reflect individual biases and abnormalities. Finally, SPs can be actively remodeled in a long-term manner by focused and intense cortical training.

Spontaneously emerging spatiotemporal neuronal activity patterns

Although most cognitive neuroscience research has traditionally focused on mapping the details of task-induced activation patterns, in recent years it has become evident that brain activity in the absence of such overt tasks is also highly informative. Indeed, following the pioneering observations of Arieli and colleagues in the visual cortex of anesthetized cats [1,2] and Biswal *et al.*'s study of human motor cortex [3], it became clear that even in the absence of a task – that is, in what appears to be a state of rest – the cerebral cortex generates rich and consistent spatiotemporal patterns of activity. In blood oxygen-level-dependent functional MRI (BOLD-fMRI), these spontaneously emerging 'resting-state' activity fluctuations (see Glossary) appear to span the entire cortical mantle and are of similar amplitude to those produced during task performance [4]. Furthermore, these activity patterns have now been documented in human single unit and local field potential (LFP) recordings as well [5–8], revealing that the dynamics of these spontaneously emerging patterns are far slower than typical task activations [7].

Despite extensive research, the functional role of spontaneous fluctuations remains elusive, although a number of hypotheses have been put forward in an attempt to explain it (Box 1) [9–11]. However, irrespective of the specific functions that spontaneous activity plays in cortical processing, one can ask what kind of information can be deciphered from these spontaneous patterns.

In what follows, extending previous notions about SPs, we propose the hypothesis – termed spontaneous trait reactivation (STR) – that SPs, at least as revealed in the human cortex, could offer a window into the structure of an individual's inner world (i.e., the unique profile of personality traits), tendencies, and even pathologies. More specifically, we propose that in the human cortex the correlation structure or functional connectivity (FC) revealed by spontaneous fluctuations is informative about

Glossary

Connectivity bias: strength or efficacy of excitatory synaptic coupling between two neurons or between neuronal groups.
Default mode/resting state/task negative network: network of brain areas, located in specific regions of medial frontal, parietal, and anterior temporal cortex, whose neuronal activity increases on entering a resting state.
Default/resting state fluctuations: ultra-slow fluctuations in neuronal activity that occur in the absence of an explicit task. These fluctuations emerge in networks across the entire cortex and are not confined to the default mode network. Note that these networks are often referred to as resting-state networks.
Functional connectivity/functional correlation: extent to which the neuronal activity across two brain sites is co-modulated. High functional connectivity may be generated by external sources (e.g., common inputs) and hence does not always imply connectivity biases.
Hebbian learning: fundamental learning rule by which inter-neuronal connectivity biases are strengthened following neuronal correlated activations and weakened if the neuronal activity is de-correlated.
Intrinsic/self-projection network: similar to the default mode network but with the added assumption that this network is specialized for internally oriented, self-related processes.
Ongoing activity: similar to the resting-state fluctuations but includes the assumption that task-related activations are superimposed on top of the spontaneous – and hence ongoing – fluctuations.
Resting state: condition in which an individual is not engaged in any focused and intentional cognitive task.
Spontaneous fluctuations: similar to the resting-state fluctuations, but can also occur during task performance in networks that are not activated by the task.
Spontaneous patterns: matrix of correlation levels generated by spontaneous fluctuations across all cortical sites.

Corresponding author: Malach, R. (rafi.malach@gmail.com).
Keywords: *a priori* biases; spontaneous activity; resting state; human; cortex; BOLD-fMRI.

1364-6613/$ – see front matter
 http://dx.doi.org/10.1016/j.tics.2013.09.014

Selective attention for modulation of inter-areal communication

Chris Lewis & Pascal Fries (ESI)

Goals

We are interested in inter-areal communication. We seek to understand how local populations across distributed brain regions organise into coordinated activity for the execution of tasks and the instantiation of coherent cognitive states. We specifically investigate selective information routing using the cognitive variable of 'attention' in order to modulate inter-areal coupling. It has been shown that attention allows enhanced representation of attended locations, features or objects at the expense of unattended locations and items. We use these top-down biases in order to dynamically modulate inter-areal coupling and measure information flow along sensory hierarchies. Such tasks are useful in order to localise the areas of the brain involved in the top-down signals required to deploy selective attention to a particular location or feature. They can also be used to reveal patterns of inter-areal coupling indicative of selective information transfer and selection.

Stimuli and Task

A number of tasks have been proposed as robust localisers for attention. In general, a cue of some sort is delivered to the subject, and based on that cue, the subject is informed what stimulus attributes are relevant for behaviour.

For example, in the case of selective spatial attention, a cue is given (either through the same modality, i.e. visual, or another modality, i.e. auditory) which instructs the subject the precise position at which a future event should be detected or discriminated. A classic case of such a task is the Posner paradigm[223], in which a visual cue is given towards one of two locations at which a future event is likely to occur. An event then occurs, while attention is deployed to that position and increased accuracy and reduced reaction times are indicative that attention has enhanced behavioural performance at that position. Similar paradigms exist for more locations, or different varieties of attention, such as feature based attention (i.e. attention to a colour, motion direction, orientation, etc)[224,225]. Further, the cues can be used either to specify a preferred response contingency, or in order to indicate the relative likelihood of a given event. In the first case, changes occurring in the non-cued location must be ignored and only cued changes should propagate and result in a response. In the second case, both cued and uncued events may lead to a response.

Main results

Execution of a task, such as the Posner task, often reveal activity across a brain-wide front-parietal network, commonly referred to as the dorsal attention network[226]. Likewise, attentional modulation is seen at all levels of the sensory hierarchy within the attended modality. We are specifically interested in the modulation of sensory areas in the selective transmission of information from discrete lower-order areas onto higher order areas[4,5]. In this case, one higher order area receives converging input from two, distinct low order areas. Attention is thought to bias the transmission of one of the lower order populations to the higher order population at the expense of the other[227].

Neural correlates of bodily self-consciousness

Nathan Faivre & Olaf Blanke (EPFL)

Goals

The main goal for this localiser 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?''). We will rely 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. 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 will rely on fMRI coupled with robotics in order to measure changes of BOLD signal associated with the FBI.

Stimuli and Task

Participants will see a virtual rod moving vertically along the midline of the virtual body's back. A custom-made robotic device will generate 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 will actuate the stimulation sphere over a rack-and-pinion mechanism. Motion will be 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 will be replaced by an object, for which no full-body illusion is expected to occur. Therefore, analyses will be performed according to a 2 x 2 factorial design, with Object (body; object) and Stroking (synchronous; asynchronous) as main factors. As an objective measure of the FBI, participants will be asked to estimate the time that a ball they hold in their hands would take to hit the ground if they were to release it, providing repeated quantifiable measurements of self-location (height above the ground). In addition, questionnaires given at the end of the experiment will serve a subjective measures. In case a robotic platform is not available, FBI can also be induced using cardio-visual stimulations, in which participants view an image of their own body (a 'virtual body') glowing either in synchrony or asynchrony with their heartbeat.

Main results

Based on previous results[13], differences of activity in synchronous vs asynchronous stroking for body vs. object are expected in bilateral temporo-parietal junctions (parietal operculum), right middle-inferior temporal cortex (including the extrastriate body area). In addition, we expect functional connectivity between the bilateral TPJ and the supplementary motor area, ventral premotor cortex, insula, intraparietal sulcus and occipitotemporal cortex[228].

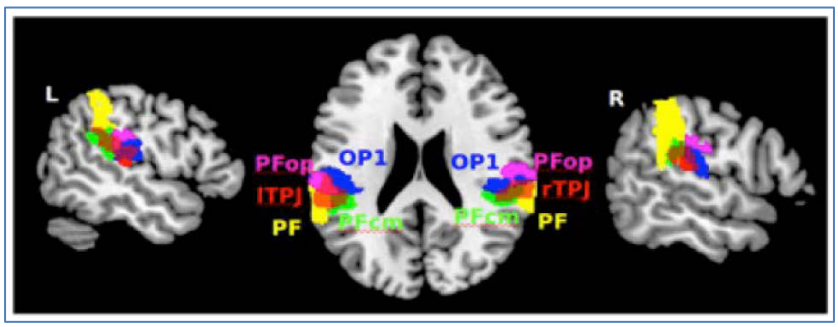


Figure 20: Anatomical overlap of regions encoding the subjective sensation of self location and first-person perspective (rTPJ & lTPJ) and cytoarchitectonically defined regions of the parietal operculum and the inferior parietal cortex

A localiser for confidence

Florent Meyniel (CEA) & Mariano Sigman (CEA/Buenos Aires)

Goals

The aim of this localiser is to dissect how confidence in a decision builds up from noisy evidence. More precisely, to identify if a neural correlate of this decision confidence can be found in the brain, and how such a neural signal is constructed from the decision processes, namely: the accumulation of evidence and the selection of a response. To decompose the decision process, we use a sensory discrimination task that is strictly controlled experimentally, in which the decision relies on the integration of pieces of evidence presented as successive images. By means of psychophysical kernels developed in a previous study, we can reconstruct how samples of evidence are combined into the response selected (right or left stimulus) and the subjective confidence level provided (high vs. low). Taking advantage on the retinotopic organisation of the early visual cortex and these psychophysical kernels, the paradigm is used 1) to track the sensory evidence at the neural level, in particular by contrasting the signals corresponding to the left and right stimuli and 2) to unravel how this activity can be read out by other brain regions to select an answer (right / left) and a confidence level.

Stimuli and Task

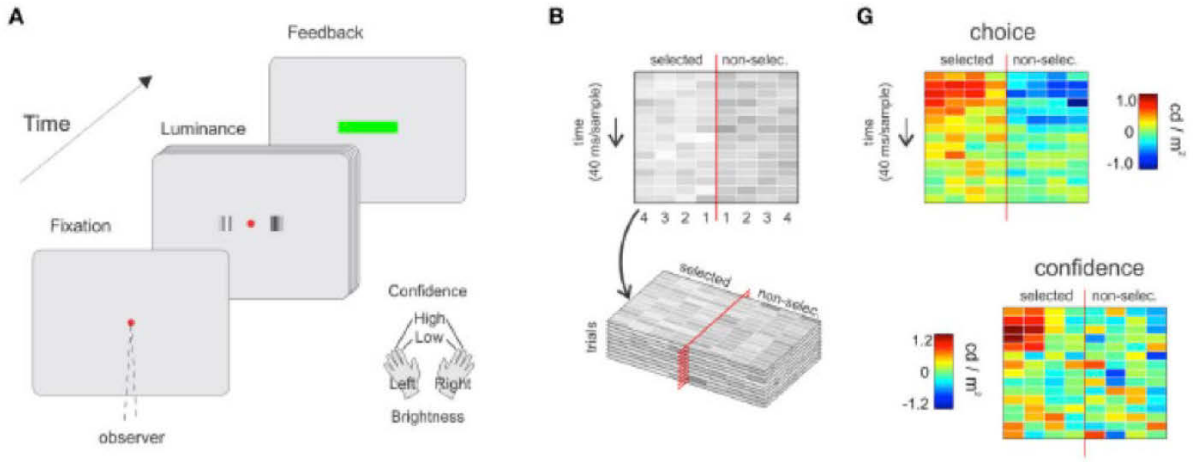


Figure 21: Localiser confidence trial structure

A trial is structured as follows: (A) Two patches of flickering bars are displayed on the left and right side of a fixation dot. The participant has to indicate the brighter patch (right or left), and his level of confidence in this response. (B) The flickering bars actually correspond to the rapid presentation of several images in which bars have different levels of luminance. Pannel (G) provides the results of the psychophysical kernels fitted given the behavioural responses to quantify the weight of each sample of evidence on the decision. Each column denotes a bar (a patch comprised 4 bars) and the rows correspond to successive images within a trial. On average across trials, the luminance of some bars were weighted more (hotter colours) or less (colder colours) in the discrimination than the level of evidence that they actually convey (i.e. their actual level of luminance). The extreme weights (hot and cold colours) on the first images reveal that the decision was based mostly on the first images, neglecting the last ones. The same kind of analysis can be applied to fit the confidence response. Note that the patterns of weights are correlated between the discrimination and confidence responses. However, this correlation is only partial, leaving open the possibility their correlates in the brain are partially distinct.

This behavioural analysis will be mirrored by a neuroimaging analysis. Using fMRI, we will quantify how the evidence provided by each sample is translated into brain responses, which in turn contribute to the selection of a response and a confidence level.

Reference

Zylberberg A, Barttfeld P, Sigman M (2012) The construction of confidence in a perceptual decision. Front Integr Neurosci (6)

Localiser – Skills and habits (procedural memory)

Avi Karni's group

Goals

The aim of these studies is to address cortical dynamics (repetition suppression, repetition enhancement, functional connectivity) as brain signatures of accumulating experience, plasticity and procedural memory consolidation in motor skill learning in typical young adults. A key strategic point is that we are following on the work of *Karni et al, 1995, 1998* by looking at short term modulations of the evoked BOLD signals in motor cortex and the motor system in general as enduring signatures of previous experience and importantly as signatures of overnight procedural memory consolidation. Thus, we are testing the conjecture that activity in a given brain area is modulated by task repetition in a differential manner as a function of whether prior experience was afforded, i.e., reflecting local mnemonic processes specifically, procedural memory consolidation. Complementing this focus on the temporal modulation of activity (as reflected in the metabolic BOLD signal) are tests for repetition dependent modulations of the functional connectivity between areas engaged in the performance of the task.

In Experiment 1 brain activation induced by actual performance are studied; in Experiment 2 we plan to compare brain activations and repetition effects evoked by actual performance to activation evoked by movement observation of the identical movement sequences to explore motor learning and memory consolidation by action observation.

Stimuli and Task

The basic study protocol is the following: On day 1 all participants are instructed (full declarative knowledge) and shown a specific sequence composed of 5 opposition movements, identical to the sequences used by *Karni et al, 1998, Korman et al, 2003, 2007*. Performance tests (speed, accuracy; video analysis) are performed before and after a structured training session that has been shown to trigger the expression of significant delayed gains in performance, overnight, in the majority of participants (Figure 22). On day 2, participants are scanned during the paced performance of either the movement sequence intensively trained a day earlier, or a similarly constructed, but novel, untrained sequence. Both movement sequences are performed in pairs of blocks separated by a brief rest interval (30 sec). The scanning session is followed by a third performance test to assess the expression of delayed "offline" gains. In Experiment 2 actual physical performance of the task is substituted (in training and in the imaging session) by observation of the target sequences (trained & untrained).

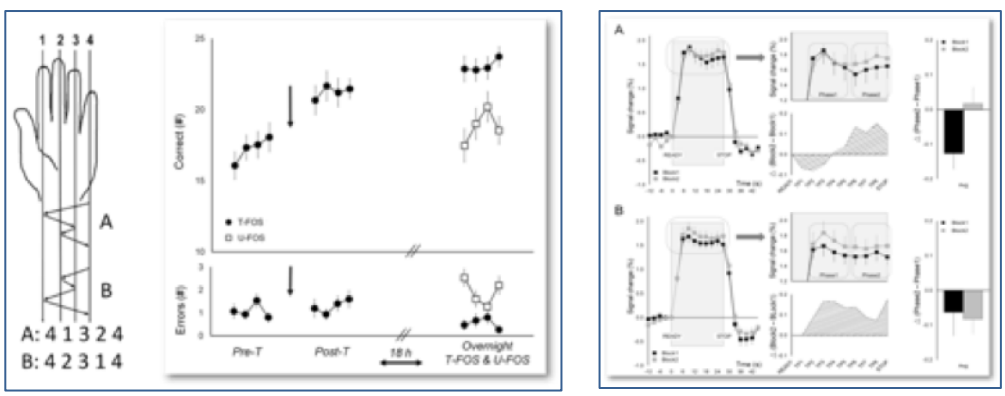


Figure 22: Procedural memory experimental results

Task. The 2 five-element finger-to-thumb opposition sequences (FOS). Each participant trained on one FOS, the other sequence served as the untrained control condition. Behavioural Results. Performance tests were conducted before and immediately after the training and at 24 hours post-training (Pre-T, Post-T, Overnight, respectively, Figure 22). arrow – training session (160 repetition of the assigned FOS). Robust gains in speed and accuracy are expressed for the trained sequence (T-FOS, solid circles). BOLD signal modulations. A. U-FOS; B. T-FOS. The relative repetition enhancements, in M1 contra-lateral to the performing hand, for the T-FOS were correlated with the overnight gains in performance (Gabitov et al, JCN, 2014).

Results

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 (Figure 23). We propose that M1 not only generates movements but also serves as a hub for a motor working memory system: wherein transient stabilisation of activity upon sequence repetition reflects short-term familiarity with a novel sequence of movements. A temporarily stabilised network in cortex and striatum may promote an integrated representation of the new movement sequence (i.e., the movement syntax). We also find that 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. A paper is currently in review (Gabitov, Manor, Karni).

Experiment 2 - behavioural (completed) and fMRI brain imaging study, to address motor cortex plasticity driven by visual input (action observation) is in progress. The behavioural data suggest that executing and observing movements improve task performance and trigger skill consolidation processes. However, consolidation could be blocked by ensuing action but not by observation, indicating that skills acquired in doing or observing do not overlap in the brain. A paper is in prep (Maaravi-Hesseg, Gal, Karni).

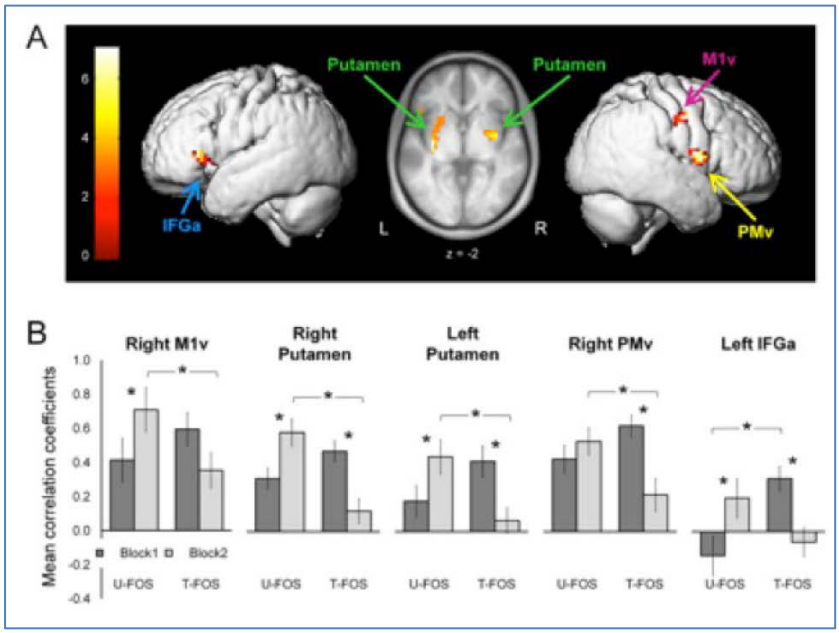


Figure 23: Functional connectivity analyses using M1 hand area as a seed

(A) Areas wherein functional connectivity patterns, were differentially modulated by the repeated performance of the two sequences using M1 as a seed (interaction: [U-FOS: Block1 < Block2] X [T-FOS: Block1 > Block2]). Connectivity map of group effects is overlaid on the mean structural image of all subjects and the surface rendered from the subjects' mean structural segmented images. The map was thresholded at $p < 0.001$. The colour-bar represents t-values. M1v – primary motor cortex ventral to the hand area, PMv – ventral premotor cortex, IFGa – anterior part of the inferior frontal gyrus. (B) Mean correlation coefficients between the M1 seed and each of the clusters, wherein connectivity with the M1 seed showed significant task by block interaction. Columns – mean Fisher-transformed correlation coefficients; bars – standard errors of means (s.e.m.). * – significant differences at 0.05 level corrected for multiple tests (i.e., a number of clusters) using Bonferroni adjustments.

Supplement :

http://www.bio-conferences.org/articles/bioconf/pdf/2011/01/bioconf_skills_00047.pdf

BIO Web of Conferences 1, 00047 (2011)
DOI: 10.1051/bioconf/20110100047

When and Where in Skill Memory Consolidation: Neuro-Behavioral Constraints on the Acquisition and Generation of Procedural Knowledge.

Avi Karni* Maria Korman†
(*)Department of Human Biology & the E.J. Safra Brain Research Center, University of Haifa, Israel;(†)Technion, Haifa, Israel
E-mail: Avi.karni@yahoo.com, korman.maria@gmail.com

Abstract

Compelling behavioral and neuro-imaging data suggest that the retention and perfection of skills (procedural, "how to" knowledge) reflects long-lasting experience-driven changes in the brain's organization (neural plasticity). Two corollaries require consideration in designing effective skill learning programs. i) Neuro-behavioral constraints, imposed on whether neuronal plasticity is triggered and allowed to proceed, must be satisfied; otherwise, the skill may fail to consolidate into long-term memory. These include the amount of task iterations afforded, task scheduling, behavioral relevancy and the degree of consistency of the to-be-learned experience over a required time-window. ii) The performance of a given task reflects qualitatively different task solution routines in different phases of experience. Practice, given time and sometimes time-in-sleep, can trigger processes whereby new procedural knowledge and qualitative changes in task solution, emerge and consolidate. These emerging changes in procedural knowledge result in differences in the ability to transfer gains, across stimulus, context and task parameters.

Our aim here is to present a very brief account of our current view of procedural learning and procedural memory consolidation as emerging from a number of studies addressing the characteristics of human skill learning. We present a number of points which we believe are of relevance to the understanding of the biological mechanisms and specifically, the neuro-behavioral constraints imposed by these mechanisms on skill learning and skill memory. These constraints require consideration if we are to improve and perhaps even optimize skill teaching protocols and skill learning programs. The references provided, mainly from our own work on perceptual learning and motor skill acquisition can be consulted for perusing the actual data and as pointers to many related studies and papers that inspired us.

A widely accepted view is that skills and habits - procedural, "how to" knowledge – are organized in memory in a manner which is quite distinct from that of declarative, "what" knowledge (knowledge of facts and singular events) and are retained in an outstandingly robust manner (Squire, 1986; Dudai, 2004). It is clear, nevertheless, that the two memory systems interact in healthy, typical adults (Albouy et al., 2008). A somewhat different perspective on the two long-term memory systems is that while the declarative memory system is geared to retain sparsely occurring and even singular experiences, the declarative memory system is triggered if and only when a given experience is repeated; preferably in a consistent manner. Thus, declarative memory can be viewed as a memory system evolved from the need to retain singular events (e.g., painful event that one would not want repeated) and arbitrary associations, while procedural memory reflects the evolving 'deep' knowledge of the structure of the relevant environment and context, as well as the statistics of repeated experiences. In the context of skill learning, a training experience will often result in some trace of declarative memory, but if training is continued, i.e., more iterative experience on the task is afforded and experience accumulates, the neural mechanisms (presumably Hebbian in character (Viana & Prisco, 1984) sub-serving the generation of long-term procedural memory can be triggered.

An important advancement in our understanding of the biology of procedural memory comes from animal and human studies that clearly indicated that the establishment of long-term procedural memory is sub-served by functional and structural changes within the brain systems involved in the performance of the task, i.e., activated by the repeated experience (Karni, 1996; Merzenich, 1998). Moreover, there is evidence suggesting that the neural changes triggered by repeated experience affect two distinct levels of brain representation. Repeated experience can result: i. in neuronal changes at local processing levels that are

Article available at http://www.bio-conferences.org or http://dx.doi.org/10.1051/bioconf/20110100047

Localiser for brain circuits that trigger consolidation of realistic episodes

Yadin Dudai's group

Goal

Identify circuits that subserve triggering of consolidation of realistic episodic memory.

Hypothesis

Realistic episodes undergo binding in a temporary episodic buffer immediately when they make sense, cued by event boundaries.

Boundary conditions

In ongoing episodes, identification of representations of posited closure cues is confounded by activation of brain circuits that encode events on the fly.

Strategy

Present brief episodes to be encoded (4-20 sec), in the form of audiovisual clips, with intercalated 'rest' (no stimulus) periods. Pursue the protocol to identify brain activation time-locked to the offset of the episode and predict subsequent memory.

Protocol

Each experiment consists of a *Study* phase in the fMRI scanner and a subsequent *Test* phase outside the scanner. The procedure is aimed to identify brain regions exhibiting delayed encoding-related activation (i.e. a higher BOLD response to subsequently remembered vs. forgotten clips, initiated at clip offset), using 8 sec audiovisual clips that were not seen before by the participants. A total of 180 clips are used, of which 160 are narrative movie clips (*Movie)* and 20 are visually scrambled clips (*Scrambled*) accompanied by non-distinctive background noises. Intercalated rest period involve 4-12 (jittered) blank screen. The fMRI data acquired during the presentation of *Movie* clips are divided into subsequently *Remembered* and *Forgotten* events. Similarly, fMRI data acquired during presentation of blank screens following the clips are divided into *R-Blank*, *F-Blank* and *S-Blank* periods (blank screens following *Remembered*, *Forgotten* and *Scrambled* clips, respectively). The conjunction contrast of *F-Blank*>baseline yielded several brain regions (combined $p<0.000025$, uncorrected, minimal cluster size of 5 contiguous functional voxels), including the right hippocampal body, bilateral hippocampus head (extending to the amygdala-hippocampal junction), right optic radiations (directly posterior to the hippocampus), bilateral dorsal caudate nucleus and bilateral posterior cerebellum. The results were replicated in several studies in our laboratory and included in several publications. The identified circuit is currently predictive only at the group level, and we are attempting to modify the protocol to achieve predictability at the individual subject level.

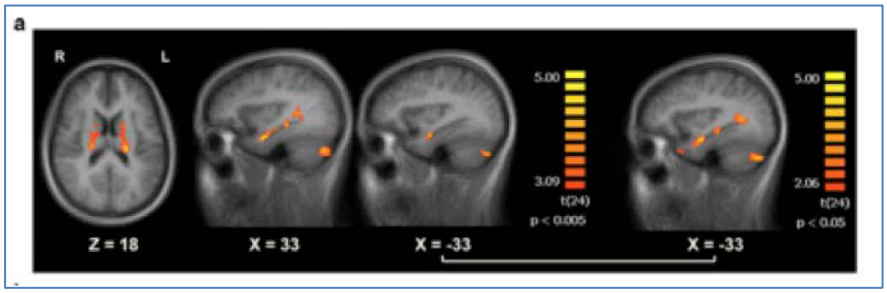


Figure 24: Brain regions predicting stimulus-offset-locked memory predicting activity, unveiled by the aforementioned protocol. Identified are striatal, hippocampal and cerebellar activations. From Ben-Yakov and Dudai 2011

References

Ben-Yakov A, Dudai Y (2011) Constructing realistic engrams: Post-stimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *J Neurosci* 31, 9032-9042; Ben-Yakov A, Eshel N, Dudai Y (2013) Hippocampal immediate post-stimulus activity in the encoding of consecutive naturalistic episodes. *J. Exp. PsychoI:G,* 142, 1255-1263; 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 (in press).*

Localiser - Working memory

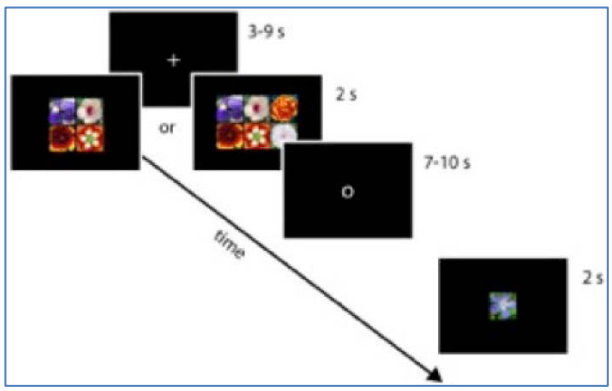
Lars Nyberg's group

Goals

The delayed match-to-sample (DMS) protocol captures the essence of WM: maintenance of information during a delay period when there is no sensory input/external support (Figure 25). Various versions of the DMS task have been used in many past human and primate studies. The proposed protocol is designed for fMRI on human participants and enables identification of brain regions that interact during temporary maintenance of information in WM, notably frontal and posterior cortical regions.

Stimuli and task

With the time parameters as in Figure 25, an experiment with 40 items takes about 13 minutes, and 30 trials will fit in 10 minutes. Should more time be allowed, possible extensions involve acquiring data on both correct and incorrect responses, and/or varying the number of items in the stimulus set so as to vary the WM load.



The different trials in the protocol are interspersed by 3-9 sec long ITIs (fixations).

1. Each trial starts by presenting a set of visual stimuli (4-6 flowers) and the participant is instructed to memorise these stimuli.

2. Then follows a delay period when the participant holds the visual information in WM for 7-10 s.

3. Lastly, the test phase consist of a match-to-sample task ("does the presented stimulus match any flower in the sample"; Yes or No?).

Figure 25: Overview of the Working Memory experimental protocol

Main results

In a pilot experiment we used 20 items of each load (in that case, 2-4 stimuli). At the individual as well as group level (18 participants), the protocol robustly elicited increased fMRI BOLD signal in several brain regions at each of the phases of the DMS task (memorisation, delay, test). Importantly, as predicted, sustained effects in fronto-temporal regions were observed during the maintenance phase (Figure 26), which is consistent with several past studies.

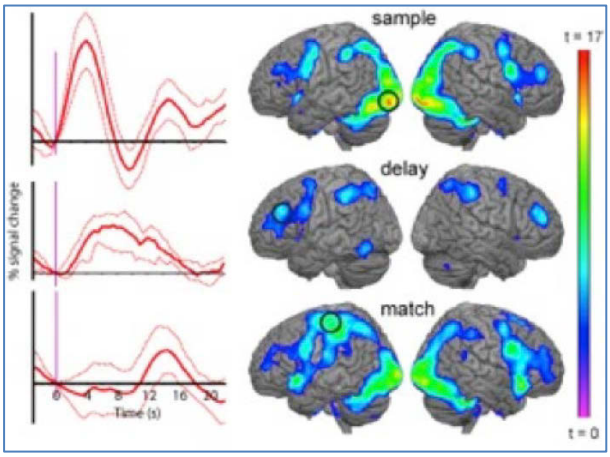


Figure 26: Working Memory - BOLD signal change for the different trial phases

Localiser for hippocampal spatial memory

Neil Burgess (UCL)

Goals

To provide a task which reliably produces detectable changes in neuroimaging (fMRI, MEG) corresponding to hippocampal processing of memory for locations within environmental space. Key issues include avoidance of confounding cognitive processes that might enable task performance without requiring knowledge of environmental location, such as visual cues that indicate the goal location directly.

Task

Desktop virtual reality provides a convenient way to probe spatial navigation within the scanner environment, and has consistently demonstrated hippocampal involvement in spatial navigation. In parallel, “subsequent memory” paradigms are the most reliable for identifying the hippocampal role in memory (i.e. correlating neural activity during encoding or planning with subsequent performance in a memory task. The Morris Water maze provides the classic task for isolating hippocampal spatial memory functions, in which there are no local cues to the goal location.

We have designed a similar virtual spatial navigation task, in which multiple object-locations have to be remembered and subsequently returned to (Figure 27). Distal cues were always present, to provide orientation, but cannot be associated with any locations as rendered at infinity. The virtual environment (a modified video game) is presented in first-person perspective on a screen; in front of the participant, who can navigate through it by pressing buttons to turn left or right or to move forwards. Participants initially collect several objects in turn from the virtual arena by navigating to them. In each subsequent trial they are cued to replace an object (the object appears on a blank background, they navigate to where they thought they had collected it and press a button). Performance is measured as replacement accuracy: the distance between the participant's response and the correct object location (positive performance is measured as the inverse of this distance). Learning trials include feedback (the object appearing in the correct location to be collected again) from which participants learn to improve performance. 'Test' trials omit this feedback. The starting location varies between trials, so that path integration/ route learning cannot be used.

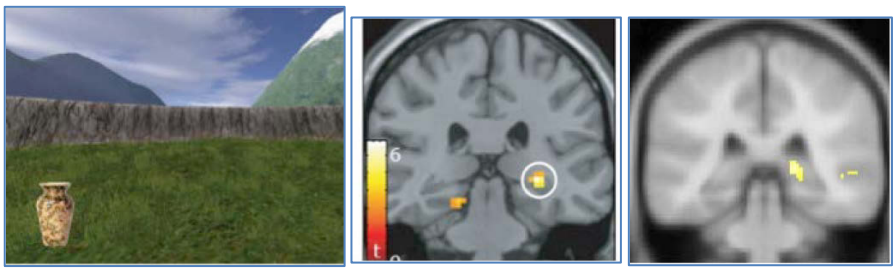


Figure 27: Virtual navigation task monitored by fMRI and MEG

Left: A view of the virtual navigation task, showing the environmental boundary (cliff), objects located within the environment (vase), and distal cues for orientation (mountains, clouds). Middle: fMRI activation in right posterior hippocampus correlates with learning object-locations[143]. Right: MEG theta power correlates with subsequent performance in remembering object-locations[229].

We have used this task to identify hippocampal fMRI activation, and MEG theta-power increases, corresponding to remembering object locations relative the environment (Figure 27).

Localiser for language comprehension

Christophe Pallier (CEA)

Goals

Language comprehension is a complex skill which involves multiple processes, notably speech perception (or letter identification in the case of reading), lexical access, syntactic parsing, semantic composition, pragmatic inference...Unsurprisingly, numerous brain areas are involved. Quasi systematically, language experiments report (relative to rest) activations in the inferior frontal and the precentral gyri, in the superior and middle temporal gyri, extending from the temporal pole to the temporo-parietal junction. In most humans, activations to language are predominant in the left hemisphere but activations in homologous regions in the right hemisphere are also typically observed, albeit to a lesser degree. The precise functional roles of each region in the language network are still a matter of active debates (which could be due to the fact that some regions are involved in different processes at different times; see Sahin et al. 2009 [230]; for attempts at systematising the findings, see the reviews by Hickok & Poeppel, 2007 [231]; Price, 2012 [232]; Friederici, 2011 [233]; Hagoort & Indefrey, 2014 [234]).

To highlight the regions involved in sentence processing in a given individual, the most common manipulation contrasts sentences to non-sentence materials. The latter can consists of scrambled lists of words, pseudowords, strings of consonants (or distorted, unintelligible speech in the case of auditory comprehension). An intermediate condition is sometimes included, which consists of syntactically correct, but meaningless materials, where contents words in sentences are replaced by pseudowords (see Indefrey et al. 2001 [235]; Humphries et al. 2006 [236]; Pallier et al. 2011 [237]).

Stimuli and Task

Three sets of materials: (1) twelve words long sentences, varying in syntactic structures and lexical content (2) matched Jabberwocky materials where content words are replaced by pseudowords having the same length (3) matched materials where all words are replaced by unpronounceable consonant strings.

These materials are presented either in the visual modality, using rapid serial visual presentation (300ms per word), or in the auditory modality (the consonant strings condition being replaced by foreign speech), using a block design (a block consisting of 3 stimuli of the same type presented in a row, separated by 1s. The block themselves are separated by 10s empty intervals). Figure 28 below shows the results of such an experiment conducted in our lab.

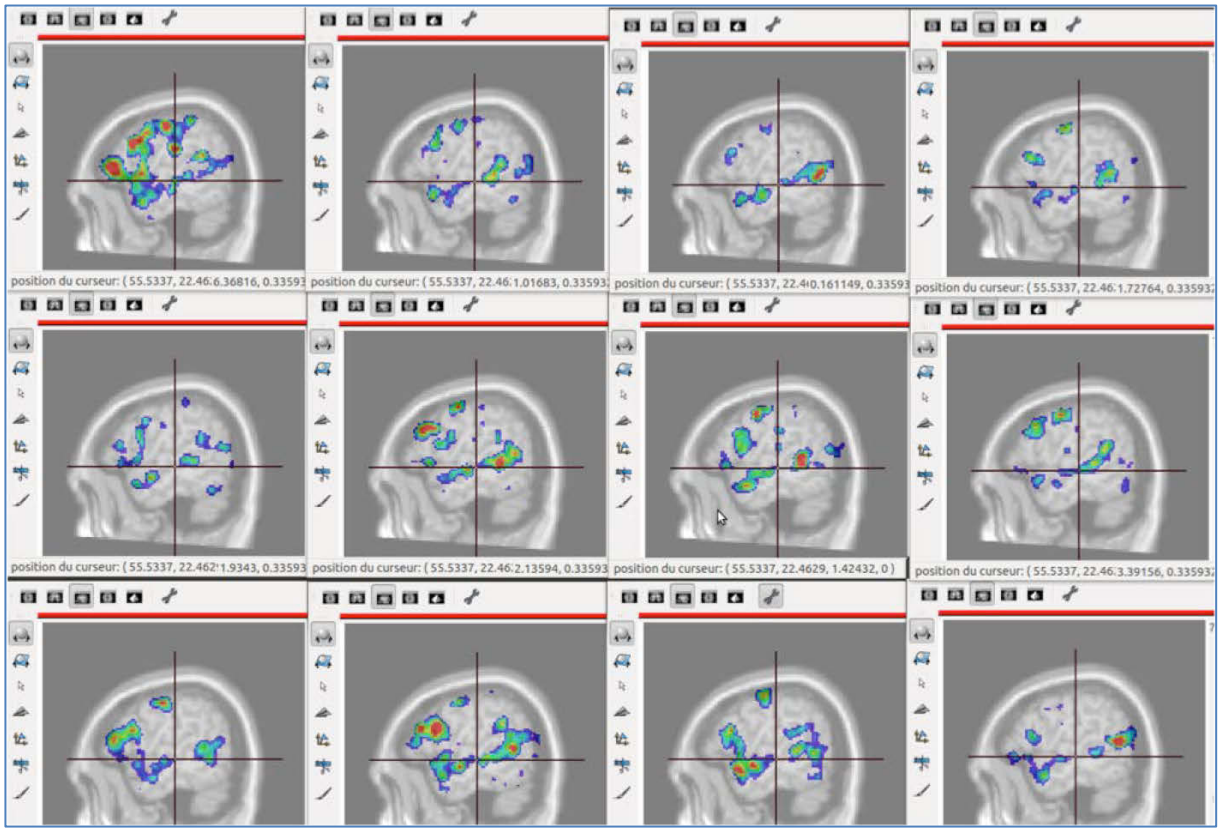


Figure 28: Activation maps obtained by contrasting sentences to consonant strings in 12 different individuals, using a 6 minutes long localiser

Social Localiser

Riitta Hari, Lauri Parkkonen, Linda Henriksson (AALTO)

Goals

We are building a localiser experiment for identifying brain regions typically associated with social cognition. Our comprehensive set of social stimuli aims to target functional subregions supporting face and body perception, action and emotion understanding, and theory-of-mind capabilities. We have the first version of the "social localiser" ready to be piloted using fMRI. In the piloting phase, we will test a large variety of stimuli and assess their effectiveness. We will then select the best stimuli to produce an efficient experimental protocol. In addition, we are considering replacing some of the well-controlled stimuli with more naturalistic social stimuli.

Stimuli and Tasks

At the moment, the localiser includes the following stimulus categories with appropriate control conditions (not listed here):

- Biological movement
- Goal-directed action perception
- Self vs. other perception
- Perception of persons in social interaction
- Theory-of-mind ability (moving geometrical shapes, false belief stories/pictures)
- Face perception (anonymous, famous)
- Body perception (body parts, whole bodies)
- Joint attention
- Emotion perception (faces, scenes)

Subjects' task is either to look at the presented pictures/videos or to respond to questions about the pictures/texts. Eye gaze will be monitored during the fMRI scanning.

Main results

The first version of the experiment is ready. We are now starting the pilot recordings and hope to have the first results by the end of the year 2014.

Annex E: References

1. Harmelech, T. & Malach, R. Neurocognitive biases and the patterns of spontaneous correlations in the human cortex. *Trends Cogn. Sci.* **17,** 606–615 (2013).

2. King, J.-R. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* **18,** 203–210 (2014).

3. Fries, P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* **9,** 474–480 (2005).

4. Bosman, C. A. *et al.* Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron* **75,** 875–888 (2012).

5. Grothe, I., Neitzel, S. D., Mandon, S. & Kreiter, A. K. Switching neuronal inputs by differential modulations of gamma-band phase-coherence. *J. Neurosci. Off. J. Soc. Neurosci.* **32,** 16172–16180 (2012).

6. Lewis, C. M., Bosman, C. A., Womelsdorf, T. & Fries, P. Stimulus induced visual cortical networks are recapitulated by local and inter-areal rhythmic synchronization. *Revis.*

7. Bastos, A. M., Vezoli, J. & Fries, P. Communication through coherence with inter-areal delays. *Curr. Opin. Neurobiol.* (2015).

8. Vanrie, J., Dekeyser, M. & Verfaillie, K. Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception* **33,** 547–560 (2004).

9. De Baene, W. & Vogels, R. Effects of adaptation on the stimulus selectivity of macaque inferior temporal spiking activity and local field potentials. *Cereb. Cortex N. Y. N 1991* **20,** 2145–2165 (2010).

10. Coombes, S., Schmidt, H. & Bojak, I. Interface dynamics in planar neural field models. *J. Math. Neurosci.* **2,** 9 (2012).

11. Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci. Off. J. Soc. Neurosci.* **16,** 2112–2126 (1996).

12. Giese, M. A. in *Artificial Neural Networks and Machine Learning – ICANN 2014* (eds. Wermter, S. et al.) 707–714 (Springer International Publishing, 2014). at <http://link.springer.com/chapter/10.1007/978-3-319-11179-7_89>

13. Ionta, S. *et al.* Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective. *Neuron* **70,** 363–374 (2011).

14. Eickhoff, S. B., Schleicher, A., Zilles, K. & Amunts, K. The human parietal operculum. I. Cytoarchitectonic mapping of subdivisions. *Cereb. Cortex N. Y. N 1991* **16,** 254–267 (2006).

15. Caspers, S. *et al.* The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual variability. *NeuroImage* **33,** 430–448 (2006).

16. Eickhoff, S. B. *et al.* Anatomical and functional connectivity of cytoarchitectonic areas within the human parietal operculum. *J. Neurosci. Off. J. Soc. Neurosci.* **30,** 6409–6421 (2010).

17. Disbrow, E., Litinas, E., Recanzone, G. H., Padberg, J. & Krubitzer, L. Cortical connections of the second somatosensory area and the parietal ventral area in macaque monkeys. *J. Comp. Neurol.* **462,** 382–399 (2003).

18. Burton, H., Sinclair, R. J. & McLaren, D. G. Cortical network for vibrotactile attention: a fMRI study. *Hum. Brain Mapp.* **29,** 207–221 (2008).

19. Burton, H., Sinclair, R. J., Wingert, J. R. & Dierker, D. L. Multiple parietal operculum subdivisions in humans: tactile activation maps. *Somatosens. Mot. Res.* **25,** 149–162 (2008).

20. Goltz, D., Pleger, B., Thiel, S., Villringer, A. & Müller, M. M. Sustained spatial attention to vibrotactile stimulation in the flutter range: relevant brain regions and their interaction. *PloS One* **8,** e84196 (2013).

21. Romo, R., Hernández, A., Zainos, A., Lemus, L. & Brody, C. D. Neuronal correlates of decision-making in secondary somatosensory cortex. *Nat. Neurosci.* **5,** 1217–1225 (2002).

22. Torquati, K. *et al.* Comparison between SI and SII responses as a function of stimulus intensity. *Neuroreport* **13,** 813–819 (2002).

23. Silani, G., Lamm, C., Ruff, C. C. & Singer, T. Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *J. Neurosci. Off. J. Soc. Neurosci.* **33,** 15466–15476 (2013).

24. Avillac, M., Denève, S., Olivier, E., Pouget, A. & Duhamel, J.-R. Reference frames for representing visual and tactile locations in parietal cortex. *Nat. Neurosci.* **8,** 941–949 (2005).

25. Pouget, A. & Sejnowski, T. J. Spatial transformations in the parietal cortex using basis functions. *J. Cogn. Neurosci.* **9,** 222–237 (1997).

26. Rizzolatti, G., Scandolara, C., Matelli, M. & Gentilucci, M. Afferent properties of periarcuate neurons in macaque monkeys. II. Visual responses. *Behav. Brain Res.* **2,** 147–163 (1981).

27. Graziano, M. S., Yap, G. S. & Gross, C. G. Coding of visual space by premotor neurons. *Science* **266,** 1054–1057 (1994).

28. Duhamel, J. R., Bremmer, F., Ben Hamed, S. & Graf, W. Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature* **389,** 845–848 (1997).

29. Schlack, A., Sterbing-D'Angelo, S. J., Hartung, K., Hoffmann, K.-P. & Bremmer, F. Multisensory space representations in the macaque ventral intraparietal area. *J. Neurosci. Off. J. Soc. Neurosci.* **25,** 4616–4625 (2005).

30. Graziano, M. S. & Gross, C. G. A bimodal map of space: somatosensory receptive fields in the macaque putamen with corresponding visual receptive fields. *Exp. Brain Res.* **97,** 96–109 (1993).

31. Jackson, P. L., Meltzoff, A. N. & Decety, J. Neural circuits involved in imitation and perspective-taking. *NeuroImage* **31,** 429–439 (2006).

32. Touzalin-Chretien, P. & Dufour, A. Motor cortex activation induced by a mirror: evidence from lateralized readiness potentials. *J. Neurophysiol.* **100,** 19–23 (2008).

33. Roberts, M. J. *et al.* Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. *Neuron* **78,** 523–36 (2013).

34. Hall, S. D. *et al.* The missing link: analogous human and primate cortical gamma oscillations. *NeuroImage* **26,** 13–7 (2005).

35. Quiroga, R. Q., Nadasdy, Z. & Ben-Shaul, Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* **16,** 1661–87 (2004).

36. Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63,** 879–88 (2009).

37. Ray, S. & Maunsell, J. H. R. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9,** e1000610 (2011).

38. Wang, X. Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. 1195–1268 (2010). doi:10.1152/physrev.00035.2008.

39. Kilpatrick, Z. P. & Ermentrout, B. Sparse gamma rhythms arising through clustering in adapting neuronal networks. *PLoS Comput. Biol.* **7,** e1002281 (2011).

40. Sclar, G., Maunsell, J. H. & Lennie, P. Coding of image contrast in central visual pathways of the macaque monkey. *Vision Res.* **30,** 1–10 (1990).

41. Contreras, D. & Palmer, L. Response to contrast of electrophysiologically defined cell classes in primary visual cortex. *J. Neurosci. Off. J. Soc. Neurosci.* **23,** 6936–45 (2003).

42. Palva, J. M. *et al.* Neuronal long-range temporal correlations and avalanche dynamics are correlated with behavioral scaling laws. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 3585–3590 (2013).

43. Sadaghiani, S., Hesselmann, G., Friston, K. J. & Kleinschmidt, A. The relation of ongoing brain activity, evoked neural responses, and cognition. *Front. Syst. Neurosci.* **4,** 20 (2010).

44. Zylberberg, A., Barttfeld, P. & Sigman, M. The construction of confidence in a perceptual decision. *Front. Integr. Neurosci.* **6,** 79 (2012).

45. Kepecs, A., Uchida, N., Zariwala, H. A. & Mainen, Z. F. Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455,** 227–231 (2008).

46. Barthelmé, S. & Mamassian, P. Evaluation of objective uncertainty in the visual system. *PLoS Comput. Biol.* **5,** e1000504 (2009).

47. Kiani, R. & Shadlen, M. N. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* **324,** 759–764 (2009).

48. Barthelmé, S. & Mamassian, P. Flexible mechanisms underlie the evaluation of visual confidence. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 20834–20839 (2010).

49. Charles, L., King, J.-R. & Dehaene, S. Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *J. Neurosci. Off. J. Soc. Neurosci.* **34,** 1158–1170 (2014).

50. Gigerenzer, G., Hoffrage, U. & Kleinbölting, H. Probabilistic mental models: a Brunswikian theory of confidence. *Psychol. Rev.* **98,** 506–528 (1991).

51. Vickers, D. & Pietsch, A. Decision making and memory: a critique of Juslin and Olsson's (1997) sampling model of sensory discrimination. *Psychol. Rev.* **108,** 789–804 (2001).

52. Juslin, P., Winman, A. & Hansson, P. The naïve intuitive statistician: a naïve sampling model of intuitive confidence intervals. *Psychol. Rev.* **114,** 678–703 (2007).

53. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10,** 1214–1221 (2007).

54. Nassar, M. R., Wilson, R. C., Heasly, B. & Gold, J. I. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J. Neurosci. Off. J. Soc. Neurosci.* **30,** 12366–12378 (2010).

55. Summerfield, C., Behrens, T. E. & Koechlin, E. Perceptual classification in a rapidly changing environment. *Neuron* **71,** 725–736 (2011).

56. Gallistel, C. R., Krishan, M., Liu, Y., Miller, R. & Latham, P. E. The perception of probability. *Psychol. Rev.* **121,** 96–123 (2014).

57. Mathys, C., Daunizeau, J., Friston, K. J. & Stephan, K. E. A bayesian foundation for individual learning under uncertainty. *Front. Hum. Neurosci.* **5,** 39 (2011).

58. Ben-Yakov, A., Eshel, N. & Dudai, Y. Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *J. Exp. Psychol. Gen.* **142,** 1255–1263 (2013).

59. Ben-Yakov, A., Rubinson, M. & Dudai, Y. Shifting Gears in Hippocampus: Temporal Dissociation between Familiarity and Novelty Signatures in a Single Event. *J. Neurosci.* **34,** 12973–12981 (2014).

60. Pine, A., Mendelsohn, A. & Dudai, Y. Unconscious learning of likes and dislikes is persistent, resilient, and reconsolidates. *Decis. Neurosci.* **5,** 1051 (2014).

61. Weigenand, A., Schellenberger Costa, M., Ngo, H.-V. V., Claussen, J. C. & Martinetz, T. Characterization of k-complexes and slow wave activity in a neural mass model. *PLoS Comput. Biol.* **10,** e1003923 (2014).

62. Burgess, N. & O'Keefe, J. Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus* **6,** 749–762 (1996).

63. Watkins, C. & Dayan, P. Q-learning. *Mach. Learn.* **8,** 279–292 (1992).

64. O'Doherty, J. *et al.* Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304,** 452–454 (2004).

65. Pearce, J. M., Roberts, A. D. & Good, M. Hippocampal lesions disrupt navigation based on cognitive maps but not heading vectors. *Nature* **396,** 75–77 (1998).

66. Squires, K. C., Wickens, C., Squires, N. K. & Donchin, E. The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science* **193,** 1142–1146 (1976).

67. Huettel, S. A., Mack, P. B. & McCarthy, G. Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. *Nat. Neurosci.* **5,** 485–490 (2002).

68. Schvaneveldt, R. W. & Chase, W. G. Sequential effects in choice reaction time. *J. Exp. Psychol.* **80,** 1–8 (1969).

69. Mars, R. B. *et al.* 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 (2008).

70. Kolossa, A., Fingscheidt, T., Wessel, K. & Kopp, B. A model-based approach to trial-by-trial p300 amplitude fluctuations. *Front. Hum. Neurosci.* **6,** 359 (2013).

71. Lieder, F., Daunizeau, J., Garrido, M. I., Friston, K. J. & Stephan, K. E. Modelling trial-by-trial changes in the mismatch negativity. *PLoS Comput. Biol.* **9,** e1002911 (2013).

72. Grill-Spector, K., Kushnir, T., Hendler, T. & Malach, R. The dynamics of object-selective activation correlate with recognition performance in humans. *Nat. Neurosci.* **3,** 837–843 (2000).

73. Del Cul, A., Baillet, S. & Dehaene, S. Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol.* **5,** e260 (2007).

74. Fisch, L. *et al.* Neural 'ignition': enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* **64,** 562–574 (2009).

75. Quiroga, R. Q., Mukamel, R., Isham, E. A., Malach, R. & Fried, I. Human single-neuron responses at the threshold of conscious recognition. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 3599–3604 (2008).

76. Amir, Y., Harel, M. & Malach, R. Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *J. Comp. Neurol.* **334,** 19–46 (1993).

77. Nir, Y., Hasson, U., Levy, I., Yeshurun, Y. & Malach, R. Widespread functional connectivity and fMRI fluctuations in human visual cortex in the absence of visual stimulation. *NeuroImage* **30,** 1313–1324 (2006).

78. Nir, Y. *et al.* Interhemispheric correlations of slow spontaneous neuronal fluctuations revealed in human sensory cortex. *Nat. Neurosci.* **11,** 1100–1108 (2008).

79. Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R. & Fried, I. Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* **322,** 96–101 (2008).

80. Schurger, A., Sitt, J. D. & Dehaene, S. An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proc. Natl. Acad. Sci. U. S. A.* **109,** E2904–2913 (2012).

81. Chater, N., Tenenbaum, J. B. & Yuille, A. Probabilistic models of cognition: conceptual foundations. *Trends Cogn. Sci.* **10,** 287–291 (2006).

82. Kepecs, A. & Mainen, Z. F. A computational framework for the study of confidence in humans and animals. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367,** 1322–1337 (2012).

83. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415,** 429–433 (2002).

84. Kording, K. P. Bayesian statistics: relevant for the brain? *Curr. Opin. Neurobiol.* **25,** 130–133 (2014).

85. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8,** 1704–1711 (2005).

86. O'Reilly, J. X., Jbabdi, S., Rushworth, M. F. S. & Behrens, T. E. J. Brain systems for probabilistic and dynamic prediction: computational specificity and integration. *PLoS Biol.* **11,** e1001662 (2013).

87. Preuschoff, K. & Bossaerts, P. Adding prediction risk to the theory of reward learning. *Ann. N. Y. Acad. Sci.* **1104,** 135–146 (2007).

88. Vossel, S. *et al.* Spatial attention, precision, and Bayesian inference: a study of saccadic response speed. *Cereb. Cortex N. Y. N 1991* **24,** 1436–1450 (2014).

89. Vul, E., Goodman, N., Griffiths, T. L. & Tenenbaum, J. B. One and done? Optimal decisions from very few samples. *Cogn. Sci.* **38,** 599–637 (2014).

90. Kolling, N., Behrens, T. E. J., Mars, R. B. & Rushworth, M. F. S. Neural Mechanisms of Foraging. *Science* **336,** 95–98 (2012).

91. Bahrami, B. *et al.* Optimally interacting minds. *Science* **329,** 1081–1085 (2010).

92. De Gardelle, V. & Mamassian, P. Does Confidence Use a Common Currency Across Two Visual Tasks? *Psychol. Sci.* **25,** 1286–1288 (2014).

93. De Martino, B., Fleming, S. M., Garrett, N. & Dolan, R. J. Confidence in value-based choice. *Nat. Neurosci.* **16,** 105–110 (2013).

94. Paul, E. J., Boomer, J., Smith, J. D. & Ashby, F. G. Information-integration category learning and the human uncertainty response. *Mem. Cognit.* **39,** 536–554 (2011).

95. Denison, S. & Xu, F. The origins of probabilistic inference in human infants. *Cognition* **130,** 335–347 (2014).

96. Téglás, E. *et al.* Pure reasoning in 12-month-old infants as probabilistic inference. *Science* **332,** 1054–1059 (2011).

97. Tobler, P. N., Christopoulos, G. I., O'Doherty, J. P., Dolan, R. J. & Schultz, W. Risk-dependent reward value signal in human prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 7185–7190 (2009).

98. Nassar, M. R. *et al.* Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* **15,** 1040–1046 (2012).

99. Van Duuren, E., Lankelma, J. & Pennartz, C. M. A. Population coding of reward magnitude in the orbitofrontal cortex of the rat. *J. Neurosci. Off. J. Soc. Neurosci.* **28,** 8590–8603 (2008).

100. Monosov, I. E. & Hikosaka, O. Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region. *Nat. Neurosci.* **16,** 756–762 (2013).

101. Christopoulos, G. I., Tobler, P. N., Bossaerts, P., Dolan, R. J. & Schultz, W. Neural correlates of value, risk, and risk aversion contributing to decision making under risk. *J. Neurosci. Off. J. Soc. Neurosci.* **29,** 12574–12583 (2009).

102. Strange, B. A., Duggins, A., Penny, W., Dolan, R. J. & Friston, K. J. Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Netw. Off. J. Int. Neural Netw. Soc.* **18,** 225–230 (2005).

103. Kahneman, D. & Tversky, A. Variants of uncertainty. *Cognition* **11,** 143–157 (1982).

104. Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T. & Miyamoto, A. Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nat. Neurosci.* **16,** 749–755 (2013).

105. Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H. & Lisanby, S. H. Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *J. Neurophysiol.* **107,** 1556–1563 (2012).

106. Charles, L., Van Opstal, F., Marti, S. & Dehaene, S. Distinct brain mechanisms for conscious versus subliminal error detection. *NeuroImage* **73,** 80–94 (2013).

107. Bach, D. R. & Dolan, R. J. Knowing how much you don't know: a neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* **13,** 572–586 (2012).

108. Payzan-LeNestour, E., Dunne, S., Bossaerts, P. & O'Doherty, J. P. The neural representation of unexpected uncertainty during value-based decision making. *Neuron* **79,** 191–201 (2013).

109. Fitzgerald, T. H. B., Seymour, B., Bach, D. R. & Dolan, R. J. Differentiable neural substrates for learned and described value and risk. *Curr. Biol. CB* **20,** 1823–1829 (2010).

110. Gläscher, J., Daw, N., Dayan, P. & O'Doherty, J. P. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66,** 585–595 (2010).

111. Acerbi, L., Vijayakumar, S. & Wolpert, D. M. On the origins of suboptimality in human probabilistic inference. *PLoS Comput. Biol.* **10,** e1003661 (2014).

112. Kahneman, D. & Tversky, A. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47,** 263 (1979).

113. Lench, H. C., Smallman, R., Darbor, K. E. & Bench, S. W. Motivated perception of probabilistic information. *Cognition* **133,** 429–442 (2014).

114. Kahneman, D. & Tversky, A. Subjective probability: A judgment of representativeness. *Cognit. Psychol.* **3,** 430–454 (1972).

115. Mamassian, P. Overconfidence in an objective anticipatory motor task. *Psychol. Sci.* **19,** 601–606 (2008).

116. Pleskac, T. J. & Busemeyer, J. R. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol. Rev.* **117,** 864–901 (2010).

117. Fetsch, C. R., Kiani, R., Newsome, W. T. & Shadlen, M. N. Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron* **83,** 797–804 (2014).

118. Yeung, N. & Summerfield, C. Metacognition in human decision-making: confidence and error monitoring. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367,** 1310–1321 (2012).

119. Churchland, A. K. & Ditterich, J. New advances in understanding decisions among multiple alternatives. *Curr. Opin. Neurobiol.* **22,** 920–926 (2012).

120. Krajbich, I. & Rangel, A. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 13852–13857 (2011).

121. Koriat, A. The self-consistency model of subjective confidence. *Psychol. Rev.* **119,** 80–113 (2012).

122. Hoyer, P. O. & Hyvärinen, A. in *Advances in Neural Information Processing Systems 15* (eds. Becker, S., Thrun, S. & Obermayer, K.) 293–300 (MIT Press, 2003). at <http://papers.nips.cc/paper/2152-interpreting-neural-response-variability-as-monte-carlo-sampling-of-the-posterior.pdf>

123. Fiser, J., Berkes, P., Orbán, G. & Lengyel, M. Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.* **14,** 119–130 (2010).

124. Krajbich, I., Armel, C. & Rangel, A. Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.* **13,** 1292–1298 (2010).

125. Peters, J. & Büchel, C. The neural mechanisms of inter-temporal decision-making: understanding variability. *Trends Cogn. Sci.* **15,** 227–239 (2011).

126. Knill, D. C. & Pouget, A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* **27,** 712–719 (2004).

127. Pouget, A., Beck, J. M., Ma, W. J. & Latham, P. E. Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* **16,** 1170–1178 (2013).

128. Deneve, S., Latham, P. E. & Pouget, A. Reading population codes: a neural implementation of ideal observers. *Nat. Neurosci.* **2,** 740–745 (1999).

129. Costello, F. & Watts, P. Surprisingly rational: probability theory plus noise explains biases in judgment. *Psychol. Rev.* **121,** 463–480 (2014).

130. Mainen, Z. F. & Kepecs, A. Neural representation of behavioral outcomes in the orbitofrontal cortex. *Curr. Opin. Neurobiol.* **19,** 84–91 (2009).

131. Kok, P., Jehee, J. F. M. & de Lange, F. P. Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* **75,** 265–270 (2012).

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

133. Mongillo, G., Barak, O. & Tsodyks, M. Synaptic Theory of Working Memory. *Science* **319,** 1543–1546 (2008).

134. Rigotti, M. *et al.* The importance of mixed selectivity in complex cognitive tasks. *Nature* **497,** 585–590 (2013).

135. Burgess, N. Spatial memory: how egocentric and allocentric combine. *Trends Cogn. Sci.* **10,** 551–557 (2006).

136. Barry, C. & Burgess, N. Neural mechanisms of self-location. *Curr. Biol. CB* **24,** R330–339 (2014).

137. Packard, M. G. & McGaugh, J. L. Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiol. Learn. Mem.* **65,** 65–72 (1996).

138. Pearce, J. M., Roberts, A. D. & Good, M. Hippocampal lesions disrupt navigation based on cognitive maps but not heading vectors. *Nature* **396,** 75–77 (1998).

139. Hamilton, D. A., Akers, K. G., Weisend, M. P. & Sutherland, R. J. How do room and apparatus cues control navigation in the Morris water task? Evidence for distinct contributions to a movement vector. *J. Exp. Psychol. Anim. Behav. Process.* **33,** 100–114 (2007).

140. Hartley, T., Maguire, E. A., Spiers, H. J. & Burgess, N. The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron* **37,** 877–888 (2003).

141. Iaria, G., Petrides, M., Dagher, A., Pike, B. & Bohbot, V. D. Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *J. Neurosci. Off. J. Soc. Neurosci.* **23,** 5945–5952 (2003).

142. Maguire, E. A. *et al.* Knowing where and getting there: a human navigation network. *Science* **280,** 921–924 (1998).

143. Doeller, C. F., King, J. A. & Burgess, N. Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 5915–5920 (2008).

144. Hull, C. L. *Principles of behavior: an introduction to behavior theory*. (Appleton-Century, 1943).

145. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55,** 189–208 (1948).

146. Rescorla, R. A. & Wagner, A. R. in *Classical Conditioning II: Current Research and Theory* 64–99 (Appleton Century Crofts, 1972).

147. Sutton, R. S. & Barto, A. G. Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* **88,** 135–170 (1981).

148. Foster, D. J., Morris, R. G. & Dayan, P. A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* **10,** 1–16 (2000).

149. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map*. (Oxford University Press, 1978).

150. Doeller, C. F. & Burgess, N. Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 5909–5914 (2008).

151. Killcross, S. & Coutureau, E. Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex N. Y. N 1991* **13,** 400–408 (2003).

152. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24,** 167–202 (2001).

153. Hartley, T., Burgess, N., Lever, C., Cacucci, F. & O'Keefe, J. Modeling place fields in terms of the cortical inputs to the hippocampus. *Hippocampus* **10,** 369–379 (2000).

154. Chersi, F. & Pezzulo, G. Using hippocampal-striatal loops for spatial navigation and goal-directed decision-making. *Cogn. Process.* **13 Suppl 1,** S125–129 (2012).

155. Dollé, L., Sheynikhovich, D., Girard, B., Chavarriaga, R. & Guillot, A. Path planning versus cue responding: a bio-inspired model of switching between navigation strategies. *Biol. Cybern.* **103,** 299–317 (2010).

156. Sheynikhovich, D., Chavarriaga, R., Strösslin, T., Arleo, A. & Gerstner, W. Is there a geometric module for spatial orientation? Insights from a rodent navigation model. *Psychol. Rev.* **116,** 540–566 (2009).

157. Balci, F., Freestone, D. & Gallistel, C. R. Risk assessment in man and mouse. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 2459–2463 (2009).

158. Wacongne, C., Changeux, J. P. & Dehaene, S. A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci* **32,** 3665–78 (2012).

159. Gavornik, J. P. & Bear, M. F. Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nat. Neurosci.* **17,** 732–737 (2014).

160. Meyer, T. & Olson, C. R. Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc Natl Acad Sci U A* **108,** 19401–6 (2011).

161. Fiorillo, C. D., Newsome, W. T. & Schultz, W. The temporal precision of reward prediction in dopamine neurons. *Nat. Neurosci.* **11,** 966–973 (2008).

162. Kraus, B. J., Robinson, R. J., White, J. A., Eichenbaum, H. & Hasselmo, M. E. Hippocampal 'time cells': time versus path integration. *Neuron* **78,** 1090–1101 (2013).

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

164. Eagleman, D. M. *et al.* Time and the brain: how subjective time relates to neural time. *J Neurosci* **25,** 10369–71 (2005).

165. Jin, D. Z., Fujii, N. & Graybiel, A. M. Neural representation of time in cortico-basal ganglia circuits. *Proc Natl Acad Sci U A* **106,** 19156–61 (2009).

166. Laje, R. & Buonomano, D. V. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* **16,** 925–933 (2013).

167. Pegado, F. *et al.* Probing the lifetimes of auditory novelty detection processes. *Neuropsychologia* **48,** 3145–54 (2010).

168. Asaad, W. F., Rainer, G. & Miller, E. K. Neural activity in the primate prefrontal cortex during associative learning [see comments]. *Neuron* **21,** 1399–407 (1998).

169. Takeuchi, D., Hirabayashi, T., Tamura, K. & Miyashita, Y. Reversal of interlaminar signal between sensory and memory processing in monkey temporal cortex. *Science* **331,** 1443–7 (2011).

170. Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I. & Miyashita, Y. Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* **401,** 699–703. (1999).

171. Orlov, T., Yakovlev, V., Hochstein, S. & Zohary, E. Macaque monkeys categorize images by their ordinal number. *Nature* **404,** 77–80 (2000).

172. Terrace, H. S., Son, L. K. & Brannon, E. M. Serial expertise of rhesus macaques. *Psychol. Sci.* **14,** 66–73 (2003).

173. Jacob, S. N. & Nieder, A. The ABC of cardinal and ordinal number representations. *Trends Cogn Sci* **12,** 41–3 (2008).

174. Ninokura, Y., Mushiake, H. & Tanji, J. Integration of temporal order and object information in the monkey lateral prefrontal cortex. *J Neurophysiol* **91,** 555–60 (2004).

175. Shima, K. & Tanji, J. Binary-coded monitoring of a behavioral sequence by cells in the pre-supplementary motor area. *J Neurosci* **26,** 2579–82 (2006).

176. Siegel, M., Warden, M. R. & Miller, E. K. Phase-dependent neuronal coding of objects in short-term memory. *Proc Natl Acad Sci U A* **106,** 21341–6 (2009).

177. Mathy, F. & Feldman, J. What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition* **122,** 346–362 (2012).

178. Miller, G. A. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Pyschological Rev.* **63,** 81–97 (1956).

179. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274,** 1926–8 (1996).

180. Graf Estes, K., Evans, J. L., Alibali, M. W. & Saffran, J. R. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol Sci* **18,** 254–60 (2007).

181. Kaminski, J., Call, J. & Fischer, J. Word learning in a domestic dog: evidence for 'fast mapping'. *Science* **304,** 1682–3 (2004).

182. Bekinschtein, T. A. *et al.* Classical conditioning in the vegetative and minimally conscious state. *Nat Neurosci* **12,** 1343–9 (2009).

183. Uhrig, L., Dehaene, S. & Jarraya, B. A hierarchy of responses to auditory regularities in the macaque brain. *J. Neurosci. Off. J. Soc. Neurosci.* **34,** 1127–1132 (2014).

184. Fujii, N. & Graybiel, A. M. Representation of action sequence boundaries by macaque prefrontal cortical neurons. *Science* **301,** 1246–9 (2003).

185. Graybiel, A. M. The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* **70,** 119–136 (1998).

186. Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S. & Dupoux, E. Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J Neurosci* **23,** 9541–6 (2003).

187. Marcus, G. F., Fernandes, K. J. & Johnson, S. P. Infant rule learning facilitated by speech. *Psychol Sci* **18,** 387–91 (2007).

188. Marcus, G. F., Vijayan, S., Bandi Rao, S. & Vishton, P. M. Rule learning by seven-month-old infants. *Science* **283,** 77–80 (1999).

189. Teichmann, M. *et al.* The role of the striatum in rule application: the model of Huntington's disease at early stage. *Brain* **128,** 1155–67 (2005).

190. Shima, K., Isoda, M., Mushiake, H. & Tanji, J. Categorization of behavioural sequences in the prefrontal cortex. *Nature* **445,** 315–8 (2007).

191. Avarguès-Weber, A., Dyer, A. G., Combe, M. & Giurfa, M. Simultaneous mastering of two abstract concepts by the miniature brain of bees. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 7481–7486 (2012).

192. Giurfa, M., Zhang, S., Jenett, A., Menzel, R. & Srinivasan, M. V. The concepts of 'sameness' and 'difference' in an insect. *Nature* **410,** 930–933 (2001).

193. Yang, C. Ontogeny and phylogeny of language. *Proc. Natl. Acad. Sci.* 201216803 (2013). doi:10.1073/pnas.1216803110

194. Frank, S. L., Bod, R. & Christiansen, M. H. How hierarchical is language use? *Proc. R. Soc. B Biol. Sci.* rspb20121741 (2012). doi:10.1098/rspb.2012.1741

195. Musso, M. *et al.* Broca's area and the language instinct. *Nat Neurosci* **6,** 774–81 (2003).

196. Pallier, C., Devauchelle, A. D. & Dehaene, S. Cortical representation of the constituent structure of sentences. *Proc Natl Acad Sci U A* **108,** 2522–7 (2011).

197. Shetreet, E., Friedmann, N. & Hadar, U. An fMRI study of syntactic layers: Sentential and lexical aspects of embedding. *NeuroImage* **48,** 707–716 (2009).

198. Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C. & Dehaene, S. A temporal bottleneck in the language comprehension network. *J. Neurosci. Off. J. Soc. Neurosci.* **32,** 9089–9102 (2012).

199. Bastiaansen, M., Magyari, L. & Hagoort, P. Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *J Cogn Neurosci* **22,** 1333–47 (2010).

200. Bahlmann, J., Schubotz, R. I. & Friederici, A. D. Hierarchical artificial grammar processing engages Broca's area. *Neuroimage* **42,** 525–34 (2008).

201. Maruyama, M., Pallier, C., Jobert, A., Sigman, M. & Dehaene, S. The cortical representation of simple mathematical expressions. *NeuroImage* **61,** 1444–1460 (2012).

202. Hauser, M. D., Chomsky, N. & Fitch, W. T. The faculty of language: what is it, who has it, and how did it evolve? *Science* **298,** 1569–79 (2002).

203. Penn, D. C., Holyoak, K. J. & Povinelli, D. J. Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav Brain Sci* **31,** 109–30; discussion 130–178 (2008).

204. Abe, K. & Watanabe, D. Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* **14,** 1067–1074 (2011).

205. Beckers, G. J. L., Bolhuis, J. J., Okanoya, K. & Berwick, R. C. Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* **23,** 139–145 (2012).

206. Bekinschtein, T. A. *et al.* Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci U A* **106,** 1672–7 (2009).

207. Stivers, T. *et al.* Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 10587–10592 (2009).

208. Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. & Rizzolatti, G. Understanding motor events: a neurophysiological study. *Exp. Brain Res.* **91,** 176–180 (1992).

209. Rizzolatti, G. & Craighero, L. The mirror-neuron system. *Annu. Rev. Neurosci.* **27,** 169–192 (2004).

210. Hari, R., Himberg, T., Nummenmaa, L., Hämäläinen, M. & Parkkonen, L. Synchrony of brains and bodies during implicit interpersonal interaction. *Trends Cogn. Sci.* **17,** 105–106 (2013).

211. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject synchronization of cortical activity during natural vision. *Science* **303,** 1634–1640 (2004).

212. Malinen, S. & Hari, R. Data-based functional template for sorting independent components of fMRI activity. *Neurosci. Res.* **71,** 369–376 (2011).

213. Stephens, G. J., Silbert, L. J. & Hasson, U. Speaker-listener neural coupling underlies successful communication. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 14425–14430 (2010).

214. Hari, R. & Kujala, M. V. Brain basis of human social interaction: from concepts to brain imaging. *Physiol. Rev.* **89,** 453–479 (2009).

215. Hari, R. *et al.* Human primary motor cortex is both activated and stabilized during observation of other person's phasic motor actions. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **369,** 20130171 (2014).

216. Yoshida, K., Saito, N., Iriki, A. & Isoda, M. Representation of others' action by neurons in monkey medial frontal cortex. *Curr. Biol. CB* **21,** 249–253 (2011).

217. Baess, P. *et al.* MEG dual scanning: a procedure to study real-time auditory interaction between two persons. *Front. Hum. Neurosci.* **6,** 83 (2012).

218. Lindeman, M., Svedholm, A. M., Riekki, T., Raij, T. & Hari, R. Is it just a brick wall or a sign from the universe? An fMRI study of supernatural believers and skeptics. *Soc. Cogn. Affect. Neurosci.* **8,** 943–949 (2013).

219. Nummenmaa, L., Glerean, E., Hari, R. & Hietanen, J. K. Bodily maps of emotions. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 646–651 (2014).

220. Nummenmaa, L. *et al.* Emotions promote social interaction by synchronizing brain activity across individuals. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 9599–9604 (2012).

221. Nummenmaa, L. *et al.* Mental action simulation synchronizes action-observation circuits across individuals. *J. Neurosci. Off. J. Soc. Neurosci.* **34,** 748–757 (2014).

222. Lahnakoski, J. M. *et al.* Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage* **100,** 316–324 (2014).

223. Posner, M. I. & Petersen, S. E. The attention system of the human brain. *Annu. Rev. Neurosci.* **13,** 25–42 (1990).

224. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3,** 201–215 (2002).

225. Maunsell, J. H. R. & Treue, S. Feature-based attention in visual cortex. *Trends Neurosci.* **29,** 317–322 (2006).

226. Corbetta, M. *et al.* A common network of functional areas for attention and eye movements. *Neuron* **21,** 761–773 (1998).

227. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18,** 193–222 (1995).

228. Ionta, S., Martuzzi, R., Salomon, R. & Blanke, O. The brain network reflecting bodily self-consciousness: a functional connectivity study. *Soc. Cogn. Affect. Neurosci.* (2014). doi:10.1093/scan/nst185

229. Kaplan, R. *et al.* Movement-related theta rhythm in humans: coordinating self-directed hippocampal learning. *PLoS Biol.* **10,** e1001267 (2012).

230. Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D. & Halgren, E. Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* **326,** 445–449 (2009).

231. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8,** 393–402 (2007).

232. Price, C. J. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage* **62,** 816–847 (2012).

233. Friederici, A. D. The brain basis of language processing: from structure to function. *Physiol. Rev.* **91,** 1357–1392 (2011).

234. Hagoort, P. & Indefrey, P. The neurobiology of language beyond single words. *Annu. Rev. Neurosci.* **37,** 347–362 (2014).

235. Indefrey, P., Hagoort, P., Herzog, H., Seitz, R. J. & Brown, C. M. Syntactic processing in left prefrontal cortex is independent of lexical meaning. *NeuroImage* **14,** 546–555 (2001).

236. Humphries, C., Binder, J. R., Medler, D. A. & Liebenthal, E. Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J. Cogn. Neurosci.* **18,** 665–679 (2006).

237. Pallier, C., Devauchelle, A.-D. & Dehaene, S. Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 2522–2527 (2011).