

Research and Perspectives in Neurosciences

György Buzsáki
Yves Christen *Editors*

Micro-, Meso- and Macro-Dynamics of the Brain

FONDATION
IPSEN
POUR LA RECHERCHE
THERAPEUTIQUE

OPEN

 Springer

Research and Perspectives in Neurosciences

More information about this series at <http://www.springer.com/series/2357>

György Buzsáki • Yves Christen
Editors

Micro-, Meso- and Macro-Dynamics of the Brain

 Springer

OPEN

Editors

György Buzsáki
The Neuroscience Institute
New York University, School
of Medicine
New York
New York, USA

Yves Christen
Fondation Ipsen
Boulogne-Billancourt, France

ISSN 0945-6082

ISSN 2196-3096 (electronic)

Research and Perspectives in Neurosciences

ISBN 978-3-319-28801-7

ISBN 978-3-319-28802-4 (eBook)

DOI 10.1007/978-3-319-28802-4

Library of Congress Control Number: 2016936685

Springer Cham Heidelberg New York Dordrecht London

© The Editor(s) (if applicable) and The Author(s) 2016. This book is published with open access at SpringerLink.com.

Open Access This book is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this book are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

This work is subject to copyright. All commercial rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Acknowledgements

The editors wish to express their gratitude to Mrs. Mary Lynn Gage for her editorial assistance and Mrs. Astrid de Gérard for the organization of the meeting.

Contents

Hippocampal Mechanisms for the Segmentation of Space by Goals and Boundaries	1
Sam McKenzie and György Buzsáki	
Cortical Evolution: Introduction to the Reptilian Cortex	23
Gilles Laurent, Julien Fournier, Mike Hemberger, Christian Müller, Robert Naumann, Janie M. Ondracek, Lorenz Pammer, Samuel Reiter, Mark Shein-Idelson, Maria Antonietta Tosches, and Tracy Yamawaki	
Flow of Information Underlying a Tactile Decision in Mice	35
Nuo Li, Zengcai V. Guo, Tsai-Wen Chen, and Karel Svoboda	
The Visual Brain: Computing Through Multiscale Complexity	43
Yves Frégnac, Julien Fournier, Florian Gérard-Mercier, Cyril Monier, Marc Pananceau, Pedro Carelli, and Xoana Troncoso	
Grid Cells and Spatial Maps in Entorhinal Cortex and Hippocampus	59
Tor Stensola and Edvard I. Moser	
The Striatum and Decision-Making Based on Value	81
Ann M. Graybiel	
Decoding the Dynamics of Conscious Perception: The Temporal Generalization Method	85
Stanislas Dehaene and Jean-Rémi King	
Sleep and Synaptic Down-Selection	99
Giulio Tononi and Chiara Cirelli	

Psyche, Signals and Systems 107
Costas A. Anastassiou and Adam S. Shai

**Federating and Integrating What We Know About the Brain at
All Scales: Computer Science Meets the Clinical Neurosciences** 157
Richard Frackowiak, Anastasia Ailamaki, and Ferath Kherif

Index 171

List of Contributors

Anastasia Ailamaki Department of Computer Science, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Costas A. Anastassiou Allen Institute, Seattle, WA, USA

György Buzsáki The Neuroscience Institute, School of Medicine, New York University, New York, NY, USA

Center for Neural Science, New York University, New York, NY, USA

Pedro Carelli Centre National de la Recherche Scientifique (CNRS-UNIC), Unité de Neurosciences, Information et Complexité (UNIC), Gif-sur-Yvette, France

Tsai-Wen Chen Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

Chiara Cirelli Department of Psychiatry, University of Wisconsin, Madison, WI, USA

Stanislas Dehaene Collège de France, Paris, France

INSERM-CEA Cognitive Neuroimaging Unit, NeuroSpin Center, Saclay, France

Julien Fournier Centre National de la Recherche Scientifique (CNRS-UNIC), Unité de Neurosciences, Information et Complexité (UNIC), Gif-sur-Yvette, France

Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Richard Frackowiak Department of Clinical Neurosciences, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

Yves Frégnac Centre National de la Recherche Scientifique (CNRS-UNIC), Unité de Neurosciences, Information et Complexité (UNIC), Gif-sur-Yvette, France

Florian Gérard-Mercier Centre National de la Recherche Scientifique (CNRS-UNIC), Unité de Neurosciences, Information et Complexité (UNIC), Gif-sur-Yvette, France

Ann M. Graybiel McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Zengcai V. Guo Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

Mike Hemberger Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Ferath Kherif Department of Clinical Neurosciences, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

Jean-Rémi King INSERM-CEA Cognitive Neuroimaging Unit, NeuroSpin Center, Saclay, France

Gilles Laurent Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Nuo Li Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

Sam McKenzie The Neuroscience Institute, School of Medicine, New York University, New York, NY, USA

Center for Neural Science, New York University, New York, NY, USA

Cyril Monier Centre National de la Recherche Scientifique (CNRS-UNIC), Unité de Neuroscience, Information et Complexité (UNIC), Gif-sur-Yvette, France

Edvard I. Moser Kavli Institute for Systems Neuroscience and Centre for Neural Computation, Norwegian University of Science and Technology, Trondheim, Norway

Christian Müller Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Robert Naumann Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Janie M. Ondracek Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Lorenz Pammer Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Marc Pananceau Centre National de la Recherche Scientifique (CNRS-UNIC), Unité de Neuroscience, Information et Complexité (UNIC), Gif-sur-Yvette, France

Samuel Reiter Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Adam S. Shai Allen Institute, Seattle, WA, USA

Department of Bioengineering, California Institute of Technology, Pasadena, CA, USA

Department of Biology, Stanford University, Palo Alto CA, USA

Mark Shein-Idelson Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Tor Stensola Kavli Institute for Systems Neuroscience and Centre for Neural Computation, Norwegian University of Science and Technology, Trondheim, Norway

Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Lisbon, Portugal.

Karel Svoboda Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

Giulio Tononi Department of Psychiatry, University of Wisconsin, Madison, WI, USA

Maria Antonietta Tosches Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Xoana Troncoso Centre National de la Recherche Scientifique (CNRS-UNIC), Unité de Neurosciences, Information et Complexité (UNIC), Gif-sur-Yvette, France

Tracy Yamawaki Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Introduction

Neural systems are characterized by wide dynamic range, robustness, plasticity, and yet stability. How these competing ingredients are amalgamated into a system in which they all ‘live’ peacefully together is a key question to address and understand in neuroscience. Neuronal firing rates, synaptic weights, and population synchrony show several orders of magnitude distribution. This skewed dynamics is supported by a neuronal substrate with equally skewed statistics from the highly skewed distribution of synapse sizes to axon diameters and to macroscopic connectivity. How these different levels of anatomical and physiological organizations interact with each other to perform effectively was the topic of a recent event organized by the Fondation Ipsen: *Colloque Médecine et Recherche* on the “Micro-, Meso- and Macro-dynamics of the brain” (Paris, April 13, 2015). The participants of this symposium addressed the issues why such a multilevel organization is needed for the brain to orchestrate perceptions, thoughts, and actions, and this volume grew out of those discussions. The individual chapters cover several fascinating facets of contemporary neuroscience from elementary computation of neurons, mesoscopic network oscillations, internally generated assembly sequences in the service of cognition, large-scale neuronal interactions within and across systems, the impact of sleep on cognition, memory, motor-sensory integration, spatial navigation, large-scale computation, and consciousness. Each of these topics requires appropriate levels of analyses with sufficiently high temporal and spatial resolution of neuronal activity in both local and global networks, supplemented by models and theories to explain how different levels of brain dynamics interact with each other and how the failure of such interactions results in neurologic and mental disease. While such complex questions cannot be answered exhaustively by a dozen or so chapters, this volume offers a nice synthesis of current thinking and work-in-progress on micro-, meso-, and macrodynamics of the brain.

New York City
Paris

György Buzsáki
Yves Christen

Hippocampal Mechanisms for the Segmentation of Space by Goals and Boundaries

Sam McKenzie and György Buzsáki

Abstract In memory, the continuous flow of experience is punctuated at meaningful boundaries between one episode and the next. When salient events are separated by increasing amounts of space or time, memory systems can accommodate in two ways. One option is to increase the amount of neural resources devoted to longer event segments. The other is to maintain the same neural resources with sacrificed spatiotemporal resolution. Here we review how the spatial coding system is affected by the segmentation of space by goals and boundaries. We argue that the resolution of the place code is dictated by the amount of space encoded within periods of theta. Thus, the theta cycle is viewed as a ‘neural word’ that segregates segments of space and its cognitive equivalents (memory, planning). In support of this conclusion, we report that, as rats traverse a linear track, the beginning of a journey is represented at the falling phase of theta whereas the journey’s end is represented on the ascending phase. The current location is represented in the temporal context of the past and future event boundaries. These results are discussed in relation to the changes in physiology observed across the longitudinal axis of the hippocampus, with a special consideration for how sequence information could be integrated by downstream ‘reader’ neurons.

Introduction

A typical morning is naturally described by a sequential list of events that are demarked by completion of sub-goals, like making a pot of coffee, leaving the apartment, and encounters with people during the subway commute. This discretization of experience has a profound influence on how information is learned and recalled (Kosslyn et al. 1974; Block 1982; Kahl et al. 1984; McNamara 1986;

S. McKenzie • G. Buzsáki (✉)

The Neuroscience Institute, School of Medicine, New York University, New York, NY 10016, USA

Center for Neural Science, New York University, New York, NY, USA

e-mail: gyorgy.buzsaki@nyumc.org

© The Author(s) 2016

G. Buzsáki, Y. Christen (eds.), *Micro-, Meso- and Macro-Dynamics of the Brain*, Research and Perspectives in Neurosciences, DOI 10.1007/978-3-319-28802-4_1

Mensink and Raaijmakers 1988; Montello 1991; Howard and Kahana 2002; Kurby and Zacks 2008; Unsworth 2008; Kiliç et al. 2013). Depending on the spacing of salient events, varying extents of space and time can be chunked together in memory. For instance, the start and end points of journeys of different length serve as salient boundaries that influence memory segmentation (Downs and Stea 1973; Golledge 1999; Bonasia et al. 2016).

Memory for events that unfold over space and time is known to depend upon the hippocampus (Tulving and Markowitsch 1998; Eichenbaum 2004; Buzsáki and Moser 2013). Recordings from hippocampal place fields have shown that salient locations and physical boundaries influence the neural representation of space. For example, when the physical size of a familiar space is extended, place field size shows a concomitant expansion (O’Keefe and Burgess 1996; Diba and Buzsáki 2008). Rescaling of the place field size has the effect of decreasing the resolution of the hippocampal code for that space. The critical role boundaries play in dictating the organization of memory may be due to an underlying influence on place field organization (Krupic et al. 2015).

Map-based spatial navigation has at least four requirements: first is the existence of a cognitive map (O’Keefe and Nadel 1978); second is self-localization on that map (O’Keefe and Nadel 1978); third is an appropriate orientation of the map assisted by the head-direction system (Ranck 1984); and fourth is the calibration of the distance scale of the map with the help of external landmarks. This latter requirement is essential for allocating neuronal resources for any journey and for an a priori determination of the place field size and their distances from each other. Currently, there is no agreed-upon mechanism to explain how the hippocampus or surrounding regions scale the representation of space.

The sequential firing of cell sequences bounded within the prominent hippocampal theta rhythm (Skaggs et al. 1996; Dragoi and Buzsáki 2006; Foster and Wilson 2007; Wang et al. 2014) may be essential for this scaling. As an extension to existing theories, we propose that the clustering of cells within theta periods defines event segmentation (Gupta et al. 2012; Wikenheiser and Redish 2015). In building this argument, we first discuss the influence that goals and landmarks have on the hippocampal representation of space. Then, we present recent electrophysiological evidence that the representations of the boundaries tend to bookend theta sequences. This observation suggests that the spatial scale of memory and the amount of allotted resources are dictated by the chunking of space within theta, which depends upon the distance between salient landmarks. Finally, we discuss outstanding challenges for sequence-based computations in the hippocampus and, potentially, other regions of the brain.

Goals and Other Boundaries Anchor and Alter the Hippocampal Place Code

Boundaries, goals and landmarks have been shown to anchor place fields (Muller et al. 1987; Knierim et al. 1995; Rivard et al. 2004). The importance of environmental geometry was clearly demonstrated in one study where rats explored a walled open arena and place fields were recorded. When rats were returned to the same space without walls, the place fields became much more diffuse and irregular (Barry et al. 2006). The walls were essential to the place field integrity. This same study found that cells that fire on one side of a boundary tend not to fire on the other, showing that spatial division causes segmentation of the hippocampal representation (Barry et al. 2006). Finally, in a study in which rats were trained to run down a linear track starting at different points, place fields tended to be anchored to either the start or end of journey (Gothard et al. 1996; Redish et al. 2000b). Fields closer to the moveable start location shifted to maintain a fixed spatial distance from the start box, whereas those fields closer to the track's end maintained their place field location even as the start box location was moved. A subset of neurons, typically with place fields in the center of the track, maintained their firing fields to the distal room cues.

These observations and others (O'Keefe and Burgess 1996) led to the hypothesis that place fields are formed by summation of input from boundary vector cells (BVCs) that fire maximally when the subject is at particular distance from a border at a preferred orientation. According to this model, hippocampal cells will fire in different locations according to the orientation and distance from a border coded by pre-synaptic neurons. In support of this model, cells that fire along boundaries have been found in the medial entorhinal cortex (mEC), the parasubiculum and the subiculum (Solstad et al. 2008; Lever et al. 2009). Importantly, if these cells fire in response to a border oriented north/south in one environment, for example, they will also fire, on the equivalent side of a parallel wall inserted in the same environment, in response to similarly oriented walls in other environments, and even to gaps that restrict movement instead of walls (Lever et al. 2009). The generality of the tuning curve suggests that the BVCs, and border cells, are truly sensitive to the edges of space.

Head direction cells that fire when subjects face a particular direction (Taube et al. 1990; Sargolini et al. 2006; Giocomo et al. 2014; Peyrache et al. 2015) may be crucial for anchoring place fields to the environmental boundaries. Consistent with this conclusion is the observation that head direction cells and place cells rotate in concert when landmarks are shifted (Knierim et al. 1995). Interestingly, head direction cells can align to different compass headings within connected regions of space (Taube and Burton 1995), further showing the critical role environmental boundaries have in segmenting the representation of space.

Another important component of the spatial coding system is the grid cells observed in mEC (Hafting et al. 2005). These cells tile the environment with multiple firing fields that are arranged in a hexagonal grid. Although the grid cell

representation was first assumed to be independent of environmental boundaries and the size of the testing arena (Fyhn et al. 2004; Hafting et al. 2005), recent grid cell studies have shown the critical role that boundaries play in dictating firing field location. In symmetrical environments, the grid appears to be aligned to the boundaries of the space (Stensola et al. 2015), whereas in non-symmetrical, open arenas grid spacing is strongly influenced by the angle at which the environmental walls meet (Krupic et al. 2015). Similarity analysis of the representation of contiguous regions of space reveals that sharp turns around corners in a zig-zag maze cause a de-correlation of the representation of neighboring spatial bins (Derdikman et al. 2009; Whitlock and Derdikman 2012). These low correlations were hypothesized to be the result of a reset of the integration of the distance travelled from the preceding wall (Derdikman et al. 2009). Similar resets have been observed in the hippocampus due to 180° turns on linear tracks (Redish et al. 2000a). Overall, these results show that the grid fields, like place fields and head direction tuning, are locked in the spatial boundaries.

In addition to walls and physical barriers, rewarded locations are also route boundaries that profoundly affect the hippocampal representation of space. Several studies have demonstrated that changing where an animal is rewarded causes cells to fire in different positions—to remap (Markus et al. 1995; Dupret et al. 2010; McKenzie et al. 2013). This remapping results in an accumulation of place fields at the goal locations (Dupret et al. 2010). Over-representation of goal locations depends upon NMDA receptor-dependent plasticity and correlates with learning (Dupret et al. 2010, 2013).

Many studies have emphasized the random nature in which place fields remap (Muller and Kubie 1987; Leutgeb et al. 2004; Vazdarjanova and Guzowski 2004; Rolls 2013; Alme et al. 2014; Rich et al. 2014). However, the remapping of place fields to goal locations can be predicted. In a recent study that addressed which cells became goal cells, rats were trained to find a new reward site in a maze in which several locations were already rewarded. Cells that began to fire at the new goal were those that had fired to other, previously learned goal locations (McKenzie et al. 2013). This distortion that reward plays on the spatial representation can be appreciated in Fig. 1. In this experiment, rats were trained to retrieve a cereal reward buried within pots that differed by how they were scented and what they were filled with (see McKenzie et al. 2014 for full details). To visualize these representations, a principal component analysis was conducted on the mean rate vectors as rats sampled each pot in each position. The first two principal components corresponded to two positions. Differences in reward potential scaled the representations along these dimensions, as if by causing a scalar increase in the firing rates of cells contributing to these components. Note that the rewarded events were associated with representations closer to the origin, due to cells that fired similarly to the rewarded item irrespective of its position (Lee et al. 2012; McKenzie et al. 2014). Therefore, the presence of reward caused some locations to be represented more similarly than others.

Grid cells, head direction cells and place cells are all anchored to boundaries and goals. In the hippocampus the presence of a goal location not only dictates where a

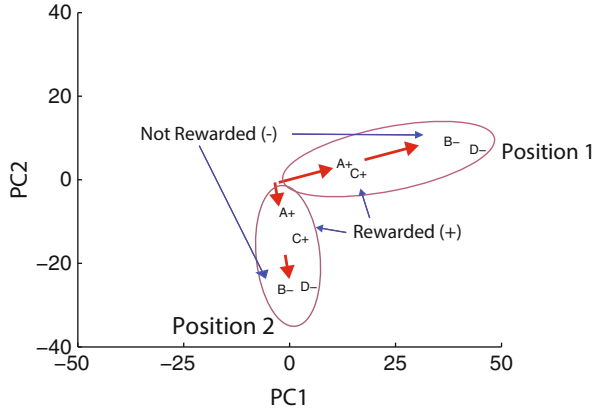


Fig. 1 Coding of rewards across different locations. CA1 and CA3 neurons ($N=438$) were recorded as rats sampled rewarded (+) and not rewarded (-) pots ($N=4$) that could appear in different positions ($N=4$). Pots differed by odor and the material in which hidden reward was buried (labeled A, B, C, D). The mean firing rate during sampling of the 16 conditions (four pots, four positions) was calculated to generate a 438×16 firing rate matrix. The first two principal components (PC) of this matrix for eight item/place combinations are plotted. The PCA was computed over all 16 item and place combinations

cell fires, but also which cells are active. In the following sections, we will argue that these salient locations anchor and distort the hippocampal spatial map by biasing which cells initiate and finish cell sequences bounded by the periods of the theta rhythm.

The Hippocampus Organizes the Spatial Code into Temporal Sequences

In addition to spatial location, hippocampal firing is modulated by the theta rhythm, which, in the rat, is a 6- to 12-Hz oscillation that can be observed in the local field potential (LFP) throughout the hippocampal system (Grastyan et al. 1959; Vanderwolf 1969; Buzsáki 2002). Early models of the origin of theta posited that hippocampal cells oscillated at theta due to an external pacemaker drive from the medial septum (Petsche et al. 1962; Lewis and Shute 1967; Lee et al. 1994). It is now clear that theta-like activity can be induced in hippocampal slices (Konopacki et al. 1988; Goutagny et al. 2009) and that there are multiple theta generators (Buzsáki et al. 1986; Kamondi et al. 1998) driven by the entorhinal cortex (Mitchell and Ranck 1980; Alonso and Llinás 1989), CA3 (Konopacki et al. 1988; Kocsis et al. 1999), the subiculum (Jackson et al. 2014), and other areas within the hippocampal circuit (Konopacki et al. 1988). Even single cells show resonance at theta frequencies (Leung and Yu 1998; Stark et al. 2013; Vaidya and Johnston 2013). Modeling work has demonstrated that a network of resonant cells can

develop rhythmic firing activity (Traub et al. 1989; White et al. 2000; Thurley et al. 2013; Tchumatchenko and Clopath 2014). Regardless of the origin of theta, the strong rhythmic activity provides temporal windows in which presynaptic inputs can be integrated, other windows in which cells fire, and windows of refractoriness in which the network is relatively silent (Buzsaki 2006).

Hippocampal pyramidal cells fire maximally at the trough of local theta (Rudell et al. 1980; Csicsvari et al. 1999). Therefore, the actual firing rate profile as subjects run through a cell's place field is a series of rhythmic bursts on a skewed Gaussian place field envelope. In a purely rate-based coding scheme, the fact that both position and theta phase dictate spiking probability presents a fundamental problem for a downstream place decoder that relies on firing rate estimation. Low firing rates could be indicative of two scenarios: either the subject is far from the center of the cell's place field, or the rat was in the center of the place field but during a non-preferred phase of theta.

Resolving this ambiguity depends upon the time scale with which presynaptic input is integrated. A systematic relationship between spiking phase and position suggests that the hippocampus is capable of sub-theta period resolution. Upon entry to the place field, cells tend to spike at late phases of theta, after the activity of the majority of other cells. Moving through the place field, not only does the firing rate increase but there is also a systematic advance in the phase in which the cell fires. In the center of the field, where firing rate is the highest, cells spike just before the chorus of other neurons. Upon exiting the field, the cell's spikes occur at early theta phases, preceding the bulk of spikes from other cells. This systematic relationship between position and the theta phase in which a cell fires is known as theta phase precession (O'Keefe and Recce 1993; Skaggs et al. 1996).

There is a close relationship between the change in rate and the change in firing phase across different types of behavior. For example, during rapid eye movement sleep, when the subject is clearly not physically moving through space, phase analysis can be done on action potentials emitted early or late in spike trains. Like in the experiments with rats running through space, spikes initiating the train are observed on late phases whereas late spikes occur on early phases (Harris et al. 2002). This phase advance can be observed in other situations. In virtual reality, phase advancement is observed in cases when spiking is fixed to virtual positions (Harvey et al. 2009; Ravassard et al. 2013) and in cases where spiking seems to occur randomly in the virtual environment (Aghajan et al. 2014). When rats run on running wheels (Harris et al. 2002; Pastalkova et al. 2008; Wang et al. 2014) or treadmills (Kraus et al. 2013), cells can become tuned to specific time intervals into running, analogous to the place field sensitivity to space. As time spent running elapses through the 'time field,' firing rates increase and decrease and precession can be observed (Pastalkova et al. 2008; Wang et al. 2014). Intriguingly, in wheel running protocols that lack a memory demand, neurons tend to fire for seconds at a fixed phase (Hirase et al. 1999; Pastalkova et al. 2008). Phase precession seems to be linked to the waxing and waning of firing rates more so than the absolute firing rate observed on a trial-to-trial basis. Phase precession is therefore a fundamental organizing principal for *changes* in the hippocampal state.

Despite decades of debate and study, there is no agreed-upon biological mechanism for phase precession. One class of model posits that phase precession is a reflection of spikes being driven by an intracellular oscillation that is a higher frequency than the theta observed in the LFP (O'Keefe and Recce 1993; O'Keefe and Burgess 2005; Hasselmo et al. 2007). Intracellular recordings of place cells show that sub-threshold oscillations increase in frequency as rats traverse a place field and that action potentials lock to peaks on this intracellular rhythm. This frequency increase rides on top of a place-locked depolarization (Harvey et al. 2009). Several models predict that that depolarization directly drives the higher frequency oscillation which causes spikes to precess (Kamondi et al. 1998; Lengyel et al. 2003). A similar conversion of a rate code into a temporal code has been suggested for spatial tuning of entorhinal grid cells, where cells are thought to integrate head direction and velocity, rather than position, into changes in firing frequency (O'Keefe and Burgess 2005; Hasselmo et al. 2007). This class of model is challenged by the observation that silencing the hippocampus for >200 ms via hippocampal commissural stimulation does not cause a reset in spiking phase (Zugaro et al. 2005). Therefore spiking phase is likely determined on a cycle-to-cycle basis, a conclusion that is at odds with oscillatory interference models for hippocampal phase precession.

A single cell mechanism of phase precession has been proposed that focuses on the rhythmic dendritic excitation that is phase synchronized with somatic inhibition (Kamondi et al. 1998; Magee 2001; Losonczy et al. 2010). Moving to the place field center causes greater amplitude dendritic excitation that progressively overcomes somatic inhibition at early times, thus causing spikes early and often (Mehta et al. 1997; Kamondi et al. 1998). This type of model requires an additional mechanism—spike adaptation (Kamondi et al. 1998) or delayed inhibition (Losonczy et al. 2010) have been suggested—to prevent late phase spiking upon place field departure. Therefore, the decrease in the period between bursts of spikes, relative to the LFP, is driven directly by increases in depolarization. Importantly, tangential passes through place fields that miss crossing the field center result in a symmetric place by position relationship, one that roughly mirrors how rate varies across position (Huxter et al. 2008). This type of behavior is predicted if peak rates must be achieved to prevent late phase spiking upon place field exit.

There have also been network models of phase precession. In a simple version of the model, CA3 place cells are wired in a feedforward chain where the cells most strongly driven fire first and at the highest rate (Jensen and Lisman 1996; Tsodyks et al. 1996; Lengyel et al. 2005). These cells then excite, through recurrent connections in CA3 and perhaps the dentate, other cells with place fields in front of the animal. Cell spiking in this model is driven both by place-related excitation and from activation of other cells with place fields between the subject and the coded position. The synaptic activation causes a time (phase) delay proportional to the distance between the current position and the cell's place field due to greater numbers of intervening cells that must be chained for more distant locations. This type of model places a heavy onus on pre-existing wiring as phase precession can be observed for first time passes through a place field (Feng et al. 2015) and in unique

trajectories in two-dimensional environments (Harris et al. 2002; Huxter et al. 2008; Jeewajee et al. 2014). These results necessitate pre-existing chains for every running direction for every position, an unlikely scenario.

In another type of network model that is not mutually exclusive with those mentioned, the most excited cells fire first, which drives inhibitory cells to delay the activity of other place cells that code for more distant positions (Dragoi and Buzsáki 2006; Maurer et al. 2006; Geisler et al. 2007; Stark et al. 2013). Silencing of soma-targeting interneurons (Royer et al. 2012), or decoupling retrograde communication between pyramidal and inhibitory cells through endocannabinoid receptor antagonism (Robbe et al. 2006; Robbe and Buzsáki 2009; Losonczy et al. 2010), causes large disruptions in assembly coordination and a redistribution of spiking across theta phase. These findings show a clear role of inhibition in phase precession.

Several observations support network models of precession. Place cells show trial-to-trial variance in their firing rates that cannot be explained by changes in position or theta phase alone (Lánský et al. 2001). Statistical models have shown that the precise trial-to-trial timing can be predicted by the spiking activity of other neurons, as would be expected if sequencing was brought about through a chaining of co-active ensembles oscillating faster than the baseline LFP (Harris et al. 2002; Dragoi and Buzsáki 2006; Geisler et al. 2007). However, principal cells respond to many environment stimuli and therefore a misspecification of the model may mistake common external modulation for a causal network interaction (Chadwick et al. 2015). Therefore further experimentation is needed to resolve whether theta sequences truly reflect network level synchronization.

Theta Sequences Code for Behaviorally Relevant Spatial Segments

Early investigators realized that phase precession could reflect cell sequences chunked into theta periods (Skaggs et al. 1996; Dragoi and Buzsáki 2006; Foster and Wilson 2007). Theta periods tend to begin with cells that have mean firing fields behind the present location and end with cells with mean fields slightly ahead. Accordingly, decoding of position on sub-theta time scales reveals spatial sequences that begin behind the animal and sweep in front (Itskov et al. 2008; Maurer et al. 2012).

Theta sequences reflect about a ten-times compression of the timing of events in the real world to time lags observed during theta (Skaggs et al. 1996) that increases with the size of the environment (Diba and Buzsáki 2008). The compression ratio can be reached by taking the cross correlation of pairs of spike trains and considering the lag in the peak at different time scales. For two place cells, the cross correlation will have a global maximum at a lag that is proportional to the distance between the place fields (Dragoi and Buzsáki 2006). These experiments are typically conducted on linear tracks with stereotyped velocity to allow a rough

equivalence between space and time. In addition, the cross correlation is strongly modulated by theta. The lags of the local maximum, on theta time scales, correlate with the time taken to traverse between the place fields. The ratio of these lags reflects the degree of compression.

A recent study explicitly tested the link between theta phase precession and theta sequencing as rats explored a novel linear track (Feng et al. 2015). This study found that phase precession was observed on the first trial, though theta sequences were not. The sequencing emerged rapidly, by the second trial, and this development coincided with a decrease in the phase variability in which cells fired upon place field entry. Therefore, theta sequencing seems to be a natural consequence of a group of cells that phase precess at the same rate (slope) and begin firing at the same phase (Dragoi and Buzsáki 2006). It is unknown what causes cells to fire at more reliable theta phases. The known importance of inhibitory cells in dictating firing phase (Royer et al. 2012) and the hypothesized role of inhibition in phase precession (Kamondi et al. 1998; Geisler et al. 2010; Losonczy et al. 2010; Stark et al. 2013) suggest a potential candidate for this phase alignment may be plasticity between excitatory and inhibitory cells. Interestingly, cells recorded at the same site tended to have more uniform phases upon place field entry (Feng et al. 2015), consistent with models in which interneurons coordinated place cells within the range of their axonal arbor.

There is growing evidence that theta sequences represent a meaningful segmentation of space. In one experiment that addressed this issue, rats were habituated to a linear track and the place field order and theta sequences were identified. Then, the track was expanded, a manipulation known to cause concomitant increases in place field size (O'Keefe and Burgess 1996). Remarkably, the theta time-scale lag remained fixed, thereby causing an increase in the compression of the amount of behavioral time represented within a theta cycle (Diba and Buzsáki 2008).

A recent experiment found that the magnitude of compression observed within each theta sequence varied significantly according to where the rat was on the maze. The amount of space represented ahead of, or behind, the rat varied systematically according to where the rat was relative to the experimentally defined landmarks (Gupta et al. 2012). This heterogeneity of theta sequence content suggests that one role of theta could be to divide space into meaningful segments.

In the aforementioned study, theta sequences could have chunked space according to the physical geometry or due to some process related to route planning. To dissociate these two possibilities, rats were trained to traverse around a circular track, collecting rewards by waiting a variable amount of time at each of three locations (Wikenheiser and Redish 2015). Rats had a choice to stay and wait for a reward or run to the next location, which was the optimal strategy if the wait time for reward at the more distant site was shorter (Wikenheiser et al. 2013). When activity on the late phases of theta was analyzed, there was a strong correlation between the distance the rat was about to run and the places represented by the active cells. Different cells spiked in the same location depending on where the rat would run next. Importantly, there was no relationship between the distance the rat had just run and the distances represented in these late theta phases. These data

showed that hippocampal activity during theta could reflect more than a representation of current state and may reflect a vicarious trial-and-error important for planning (Schmidt et al. 2013).

A similar observation has been made by decoding position using CA3 firing rates at the choice points. This analysis reveals transient moments in which CA3 represented positions ahead of the rat, sweeping down the potential paths before the rat made its decision (Johnson and Redish 2007). These findings are closely related to the fact that the phase of spiking contains information about heading direction in two-dimensional environments (Huxter et al. 2008), as would be expected if theta sequences code for upcoming positions.

Overall, studies to date have demonstrated that theta sequences always begin with place representations behind the subject and end with representations of the future. However, the exact span coded by theta sequences has not been addressed carefully. If the cells that are active at the trough of the CA1 theta cycle code for the current position in the context of past and future locations, how is the span of the past and future determined at the physiological level? One possibility is that theta sequences code for a fixed amount of time or distance around the current location. Alternatively, each geometric segment (e.g., individual corridors) and event along the journey could be represented separately as a ‘neural word’ and such words would be concatenated, perhaps via sharp wave ripples (Foster and Wilson 2006; Davidson et al. 2009; Wu and Foster 2014), to represent the entire journey from the beginning to the end. Yet another possibility is that the start and end (reward) locations of a complex trajectory through a maze are coded in a given cycle. This final option raises the question of just how much space could be segmented within a theta cycle.

Data collected in our lab demonstrate that theta periods segment the environment either according to goals or to environmental geometry. As a rat ran down the track, the probability that it occupied any given position given the observed CA1 spiking pattern was computed by comparing the instantaneous rates to a template of the session averages, the cells’ place fields. When these posterior probability distributions were calculated at every theta phase (Zhang et al. 1998), we observed theta sequences that started at one end of the track and finished at the other (Fig. 2). Thus, in addition to the goal being represented at late theta phases (Wikenheiser and Redish 2015), our findings show that the start location is represented at early phases. Combining these observations, the phase code is defined by the current location in the context of a past bounded by a journey’s beginning and a future bounded by the journey’s end. Separation of the future and past boundaries is assured by the strongest inhibition at the peak of the theta cycle (Buzsáki 2002).

Recall the studies in which place fields expanded when familiar environments were stretched. How do place fields expand with the environment? An answer begins to emerge when one considers that the theta sequences are anchored to the boundaries. The amount of space represented within the sequence, the compression, dictates the resolution of the spatial code. When boundaries are moved apart, either in the stretched environments or for journeys of different lengths, theta sequences that are bookended by those boundaries necessarily represent more space which, in

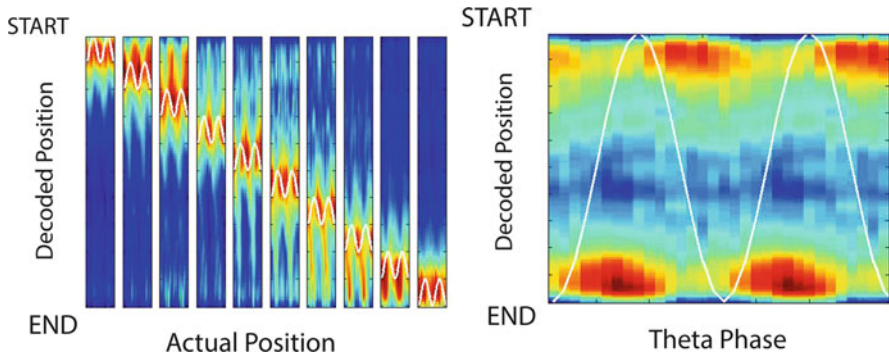


Fig. 2 *Left*, as rats run on a 1.2-m linear track, the decoded probability (high probability = red) of the rat occupying each track position (y-axis) is calculated at each phase of theta (x-axis, white sine wave). In each subplot, the range of the white sine wave demarks the rat's actual position. Generally, there is a high probability of the rat occupying its actual position. However, within a subplot, theta sequencing can be visualized by diagonal streaks of high probability that begin at the START position on the falling phase of theta and finish at the END position at the rising phase. *Right*, the same data averaged across all positions actually occupied by the rat. Note that theta sequences are bookended by representations of the linear track START and END positions at the falling and rising phases, respectively. Note that decoding was done on simultaneous ensembles measured across 4 mm of the hippocampus

turn, causes place fields to expand (Diba and Buzsáki 2008). As Redish and colleagues have shown, subjects can, on a moment-to-moment basis, allocate computational resources as a function of the planned trajectory length. Long trajectories were associated with larger place fields, and thus the resolution of the spatial code for these trials was coarser (Wikenheiser and Redish 2015). Our findings expand on these observations by demonstrating that it is not only the goal but both the beginning and end of a continuous stretch (such as a linear track) that are simultaneously represented by the theta assemblies. In our linear track experiments, the environmental boundaries and the goal locations were the same, and therefore further studies are needed to determine whether the route boundaries or the environmental geometry dictated the reliable phase coding of the start and stop locations.

Given the rapid formation of place fields upon entry into a new environment (Frank 2004; Dragoi and Tonegawa 2011; Feng et al. 2015), there must be some mechanism that estimates the spatiotemporal extent of the event segment to allocate resources appropriately. The fact that nearby neurons exhibit similarly sized place fields (Jung et al. 1994; Kjelstrup et al. 2008) suggests that there is a characteristic segment size for a species that moves through space at a particular rate. It is possible that salient events tend to happen at regular temporal or spatial intervals (Sreekumar et al. 2014). Alternatively, the segment size may depend upon internal limitations of hippocampal processing, for example, the limited amount of time in which information can be held across a delay or a limited amount of time a cell can fire at a faster rate than the overall population (Geisler et al. 2010). It is telling that,

even in large stretches of ‘open space,’ rodents choose certain spots as ‘home bases’ (Eilam and Golani 1989), perhaps to subdivide the space into spatial segments tailored for hippocampal processing. A recent study of neurons in the ventral hippocampus showed that, with learning, place fields shrank to encompass the space that equivalently predicted which objects contained a hidden reward (Komorowski et al. 2013). In this study, the default place field size was a poor predictor of the spatial extent of the context boundaries and therefore the system was modified to resolve the mismatch.

In an intriguing parallel to the organization of theta sequences, firing of cells in the ventral striatum has been shown to phase precess relative to hippocampal theta (van der Meer and Redish 2011; Malhotra et al. 2012). Cells in the ventral striatum showed ramped firing as subjects ran towards goals. Remarkably, striatal phase precession occurred over a long spatial extent for distant goals and over much shorter spatial segments when goals were close together. The phase precession appeared to be bookended by experimentally defined boundaries—the goal sites. Striatal activity might be driven by cells in the ventral hippocampus, which showed precession (Kjelstrup et al. 2008), ramped firing towards goals (Royer et al. 2010) and connectivity with the ventral striatum (Groenewegen et al. 1987). These results suggest that downstream areas may be sensitive to how space is segmented by hippocampal theta sequences (Pezzulo et al. 2014), though future studies in which both regions are recorded simultaneously are needed to assure the link between these two observations.

How Could Theta Sequences Be Integrated by Post-synaptic Readers?

Aside from the distance between place fields, there are other factors that influence the temporal lags in cell activity. The mutual dependency of the distance between place fields and anything else in determining spiking phase lag seriously complicates the aforementioned models for the computation role of cell sequences.

Cells recorded in different regions of the hippocampus have different properties. Septal CA1 cells tend to have smaller, unimodal place fields whereas more temporal cells have larger, multi-modal fields (Jung et al. 1994; Kjelstrup et al. 2008; Royer et al. 2010; Komorowski et al. 2013). Hippocampal place cells have been shown to phase precess, with spikes initiating the spike train emitted on the late phases of local theta (Maurer et al. 2005; Kjelstrup et al. 2008). Therefore, considering a pair of cells with their place fields centered at the same location, the timing difference between spikes will change in sign as the rat crosses the place fields’ common center. A range of place field sizes will cause a range in timing offsets, all of which equivalently code for the same position.

The situation is complicated further by the systematic shift in theta phase across the longitudinal axis of the hippocampus. Simultaneous recording of the LFP or

current source density analysis has shown that theta is a traveling wave (Lubenov and Siapas 2009; Patel et al. 2012) that begins at the most septal end of the hippocampus closest to the subiculum and moves temporally and proximally, resulting in a 180° phase shift at the two poles of the hippocampus (Patel et al. 2012). The speed of the travelling wave and, therefore, the maximal phase offset also change between waking and REM sleep (Patel et al. 2012). Importantly, the phase preference for spiking, with respect to local theta, does not change across the longitudinal axis (Patel et al. 2012) and, as mentioned, the phase onset and offset of precession are the same regardless of cell location (Maurer et al. 2005; Kjelstrup et al. 2008; Patel et al. 2012). Therefore, every instant in time is associated with cells at different parts of their phase precession cycle.

This observation led to the realization that moments in time do not represent points in space but could instead represent line segments (Lubenov and Siapas 2009). Since there are a range of phases that can be observed in any snapshot of time, there could theoretically be a range of represented positions, if spike phase codes for a point in space. Unless cells had equivalent place fields and were located at the same transverse lamellae along the longitudinal axis, the time delays between cells would not convey any reliable information about the distance between the place fields. The reports for this correlation in the literature are likely due to the sampling from ensembles that conform to these restrictions (Dragoi and Buzsáki 2006; Feng et al. 2015).

It is unknown whether the hippocampus acts as a single computational unit or whether transverse lamellae have different, and independent, computational roles (Andersen et al. 2000; Strange et al. 2014). If lamellae have a relative degree of independence, then the conditions could be met for phase lags to represent place field separation. Early track tracing studies showed mainly parallel fibers along transverse lamellae, implying that the trisynaptic loop is the fundamental processing module that repeats across the longitudinal axis (Andersen et al. 1969, 2000; Tamamaki and Nojyo 1991). Subsequent cell tracing studies revealed that the Schaffer collateral fans broadly from CA3 to CA1, thus allowing for substantial integration across the longitudinal axis (Amaral and Witter 1989; Ishizuka et al. 1990; Li et al. 1994), in addition to the well-known CA3 recurrent collaterals (Lorente De Nó 1934; Wittner et al. 2007). Furthermore, the axonal arborization of GABAergic cells can innervate as much 800 µm of the longitudinal axis, allowing for considerable inter-laminar crosstalk (Sik et al. 1995; see also Sloviter and Lømo 2012).

Despite this newer anatomical evidence, others have argued for relative independence of the transverse lamellae (Sloviter and Lømo 2012). Stimulation of a small region of CA3 causes maximal axonal volleys in CA1 regions in the same transverse plane (Andersen et al. 2000). Lesion and inactivation studies have also shown dissociations in the function of the septal and temporal hippocampus. Lesions to the septal hippocampus cause spatial memory deficits whereas those to the temporal hippocampus are often associated with anxiolytic measures and motivation (Moser et al. 1995; Kjelstrup et al. 2002; Pentkowski et al. 2006; Bast et al. 2009; Jarrard et al. 2012; Kheirbek et al. 2013; Wu and Hen 2014). There are

also large differences in efferent and afferent connections as well as sharp genetic variations that delineate regions across the longitudinal axis (reviewed in Strange et al. 2014). In further support of anatomical segregation is the finding that place cells in the septal versus the temporal hippocampus have been shown to remap at different rates (Komorowski et al. 2013) and to possess different place field properties on the radial arm maze, linear track, and zig-zag maze (Royer et al. 2010).

The anatomy and physiology of CA1 projections to the subiculum strongly suggest that single subicular cells have access to a large range of the longitudinal axis of CA1. Cell reconstruction studies have shown that CA1 cells project to “slabs” of the subiculum that span a narrow range of the transverse axis but up to 2 mm along the longitudinal axis (Tamamaki and Nojyo 1990, 1991). Those subicular cells would integrate across a broad range of hippocampal theta phases ($\sim 60^\circ$). *In vitro* comparisons of physiology in hippocampal slices versus that in an intact preparation showed large differences in the theta phase offsets between CA3 and the subiculum and in the theta frequency, suggesting that the slice preparation severed processes necessary for communication across lamellae (Jackson et al. 2014). Physiological studies, like those done between CA3 and CA1 (Andersen et al. 2000), are needed to determine the strength of these cross-laminar projections.

If cross-laminar communication is substantial, the compression that had been hypothesized to occur over time may occur instead over co-active neurons firing at different local phases (Lubenov and Siapas 2009). In this scheme, information is communicated by which neurons are co-active and not by their inter-spike intervals (Harris 2005). Segmentation of the environment would still be evidenced by which regions of space were represented by the ensemble at each phase, though these segments may not change within a theta period (for a different perspective see Shankar and Howard 2015).

Conclusion

Goal locations have been shown to discretize memory and to segment the hippocampal representation of space. Here we have presented evidence that salient boundaries play an important role in defining how theta sequences begin and end. We propose that this segmentation anchors place cell firing and consequently the organization of memory. However, basic questions remain as to how the hippocampal spatial code becomes coordinated during theta. What causes different areas of space to be chunked within a theta sequence and consequently the resolution of the spatial code? How do certain locations become over-represented? Are these phenomena related? How does a planning-related signal shift the represented position further ahead (or behind) the rat with the expected (or realized) journey length? Simultaneous recordings from across the longitudinal axis of the hippocampus and between the hippocampus and its output regions will help resolve the

spatial and temporal scale in which information is integrated. Finally, behavioral tests combined with recordings are required to establish whether segmentation of space into theta sequences is linked to how subjects behaviorally segment experience. Future experiments that address these questions may reveal important evidence as to how the continuous nature of experience becomes discretized in our memory.

Acknowledgments This work was supported by National Institutes of Health Grants (MH54671; MH102840), the Mather's Foundation and the National Science Foundation (Temporal Dynamics of Learning Center Grant SBE 0542013). Data presented in Fig. 1 were recorded in Howard Eichenbaum's lab. We thank Jagdish Patel for providing data recorded on the linear track for the analysis reported in Fig. 2.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Aghajian ZM, Acharya L, Moore JJ, Cushman JD, Vuong C, Mehta MR (2014) Impaired spatial selectivity and intact phase precession in two-dimensional virtual reality. *Nat Neurosci* 18:121–128
- Alme CB, Miao C, Jezek K, Treves A, Moser EI, Moser M-B (2014) Place cells in the hippocampus: eleven maps for eleven rooms. *Proc Natl Acad Sci U S A* 111:201421056
- Alonso A, Llinás RR (1989) Subthreshold Na⁺-dependent theta-like rhythmicity in stellate cells of entorhinal cortex layer II. *Nature* 342:175–177
- Amaral DG, Witter MP (1989) The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience* 31:571–591
- Andersen P, Bliss TV, Lomo T, Olsen LI, Skrede KK (1969) Lamellar organization of hippocampal excitatory pathways. *Acta Physiol Scand* 76:4A–5A
- Andersen P, Soleng AF, Raastad M (2000) The hippocampal lamella hypothesis revisited. *Brain Res* 886:165–171
- Barry C, Lever C, Hayman R, Hartley T, Burton S, O'Keefe J, Jeffery K, Burgess N (2006) The boundary vector cell model of place cell firing and spatial memory. *Rev Neurosci* 17:71–97
- Bast T, Wilson IA, Witter MP, Morris RGM (2009) From rapid place learning to behavioral performance: a key role for the intermediate hippocampus. *PLoS Biol* 7:e1000089
- Block RA (1982) Temporal judgments and contextual change. *J Exp Psychol Learn Mem Cogn* 8:530–544
- Bonasia K, Blommestejn J, Moscovitch M (2016) Memory and navigation: compression of space varies with route length and turns. *Hippocampus* 26:9–12
- Buzsáki G (2002) Theta oscillations in the hippocampus. *Neuron* 33:325–340
- Buzsáki G (2006) *Rhythms of the brain*. Oxford University Press, Oxford

- Buzsáki G, Moser EI (2013) Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nat Neurosci* 16:130–138
- Buzsáki G, Czopf J, Kondákor I, Kellényi L (1986) Laminar distribution of hippocampal rhythmic slow activity (RSA) in the behaving rat: current-source density analysis, effects of urethane and atropine. *Brain Res* 365:125–137
- Chadwick A, van Rossum MCW, Nolan MF (2015) Independent theta phase coding accounts for CA1 population sequences and enables flexible remapping. *Elife* 4
- Csicsvari J, Hirase H, Czurkó A, Mamiya A, Buzsáki G (1999) Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *J Neurosci* 19:274–287
- Davidson TJ, Kloosterman F, Wilson MA (2009) Hippocampal replay of extended experience. *Neuron* 63:497–507
- Derdikman D, Whitlock JR, Tsao A, Fyhn M, Hafting T, Moser M-B, Moser EI (2009) Fragmentation of grid cell maps in a multicompartiment environment. *Nat Neurosci* 12:1325–1332
- Diba K, Buzsáki G (2008) Hippocampal network dynamics constrain the time lag between pyramidal cells across modified environments. *J Neurosci* 28:13448–13456
- Downs R, Stea D (1973) Image and environment: cognitive mapping and spatial behavior. Transaction Publishers, Piscataway, NJ
- Dragoi G, Buzsáki G (2006) Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron* 50:145–157
- Dragoi G, Tonegawa S (2011) Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature* 469:397–401
- Dupret D, O'Neill J, Pleydell-Bouverie B, Csicsvari J (2010) The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nat Neurosci* 13:995–1002
- Dupret D, O'Neill J, Csicsvari J (2013) Dynamic reconfiguration of hippocampal interneuron circuits during spatial learning. *Neuron* 78:166–180
- Eichenbaum H (2004) Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron* 44:109–120
- Eilam D, Golani I (1989) Home base behavior of rats (*Rattus norvegicus*) exploring a novel environment. *Behav Brain Res* 34:199–211
- Feng T, Silva D, Foster DJ (2015) Dissociation between the experience-dependent development of hippocampal theta sequences and single-trial phase precession. *J Neurosci* 35:4890–4902
- Foster DJ, Wilson MA (2006) Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* 440:680–683
- Foster DJ, Wilson MA (2007) Hippocampal theta sequences. *Hippocampus* 17:1093–1099
- Frank LM (2004) Hippocampal plasticity across multiple days of exposure to novel environments. *J Neurosci* 24:7681–7689
- Fyhn M, Molden S, Witter MP, Moser EI, Moser M-B (2004) Spatial representation in the entorhinal cortex. *Science* 305:1258–1264
- Geisler C, Robbe D, Zugaro M, Sirota A, Buzsáki G (2007) Hippocampal place cell assemblies are speed-controlled oscillators. *Proc Natl Acad Sci USA* 104:8149–8154
- Geisler C, Diba K, Pastalkova E, Mizuseki K, Royer S, Buzsáki G (2010) Temporal delays among place cells determine the frequency of population theta oscillations in the hippocampus. *Proc Natl Acad Sci USA* 107:7957–7962
- Giocomo LM, Stensola T, Bonnevie T, Van Cauter T, Moser M-B, Moser EI (2014) Topography of head direction cells in medial entorhinal cortex. *Curr Biol* 24:252–262
- Golledge R (1999) Human wayfinding and cognitive maps. In: Golledge R (ed) Wayfinding behavior cognitive mapping and other spatial processes. Johns Hopkins University Press, Baltimore, MD, pp 5–45
- Gothard KM, Skaggs WE, Moore KM, McNaughton BL (1996) Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *J Neurosci* 16:823–835
- Goutagny R, Jackson J, Williams S (2009) Self-generated theta oscillations in the hippocampus. *Nat Neurosci* 12:1491–1493
- Grastyan E, Lissak K, Madarasz I, Donhoff H (1959) Hippocampal electrical activity during the development of conditioned reflexes. *Electroencephalogr Clin Neurophysiol* 11:409–430

- Groenewegen HJ, Vermeulen-Van der Zee E, te Kortschot A, Witter MP (1987) Organization of the projections from the subiculum to the ventral striatum in the rat. A study using anterograde transport of Phaseolus vulgaris leucoagglutinin. *Neuroscience* 23:103–120
- Gupta AS, van der Meer MAA, Touretzky DS, Redish AD (2012) Segmentation of spatial experience by hippocampal θ sequences. *Nat Neurosci* 15:1032–1039
- Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI (2005) Microstructure of a spatial map in the entorhinal cortex. *Nature* 436:801–806
- Harris KD (2005) Neural signatures of cell assembly organization. *Nat Rev Neurosci* 6:399–407
- Harris KD, Henze DA, Hirase H, Leinekugel X, Dragoi G, Czurkó A, Buzsáki G (2002) Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature* 417:738–741
- Harvey CD, Collman F, Dombeck DA, Tank DW (2009) Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* 461:941–946
- Hasselmo ME, Giocomo LM, Zilli EA (2007) Grid cell firing may arise from interference of theta frequency membrane potential oscillations in single neurons. *Hippocampus* 17:1252–1271
- Hirase H, Czurkó A, Csicsvari J, Buzsáki G (1999) Firing rate and theta-phase coding by hippocampal pyramidal neurons during “space clamping”. *Eur J Neurosci* 11:4373–4380
- Howard MW, Kahana MJ (2002) A distributed representation of temporal context. *J Math Psychol* 46:269–299
- Huxter JR, Senior TJ, Allen K, Csicsvari J (2008) Theta phase-specific codes for two-dimensional position, trajectory and heading in the hippocampus. *Nat Neurosci* 11:587–594
- Ishizuka N, Weber J, Amaral DG (1990) Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *J Comp Neurol* 295:580–623
- Itskov V, Pastalkova E, Mizuseki K, Buzsáki G, Harris KD (2008) Theta-mediated dynamics of spatial information in hippocampus. *J Neurosci* 28:5959–5964
- Jackson J, Amilhon B, Goutagny R, Bott J-B, Manseau F, Kortleven C, Bressler SL, Williams S (2014) Reversal of theta rhythm flow through intact hippocampal circuits. *Nat Neurosci* 17:1362–1370
- Jarrard LE, Luu LP, Davidson TL (2012) A study of hippocampal structure-function relations along the septo-temporal axis. *Hippocampus* 22:680–692
- Jeewajee A, Barry C, Douchamps V, Manson D, Lever C, Burgess N (2014) Theta phase precession of grid and place cell firing in open environments. *Philos Trans R Soc Lond B Biol Sci* 369:20120532
- Jensen O, Lisman JE (1996) Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells. *Learn Mem* 3:279–287
- Johnson A, Redish AD (2007) Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J Neurosci* 27:12176–12189
- Jung MW, Wiener SI, McNaughton BL (1994) Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *J Neurosci* 14:7347–7356
- Kahl HB, Herman JF, Klein CA (1984) Distance distortions in children’s cognitive maps: an examination of the information storage model. *J Exp Child Psychol* 38:134–146
- Kamondi A, Acsády L, Wang XJ, Buzsáki G (1998) Theta oscillations in somata and dendrites of hippocampal pyramidal cells in vivo: activity-dependent phase-precession of action potentials. *Hippocampus* 8:244–261
- Kheirbek MA, Drew LJ, Burghardt NS, Costantini DO, Tannenholz L, Ahmari SE, Zeng H, Fenton AA, Hen R (2013) Differential control of learning and anxiety along the dorsoventral axis of the dentate gyrus. *Neuron* 77:955–968
- Kiliç A, Criss AH, Howard MW (2013) A causal contiguity effect that persists across time scales. *J Exp Psychol Learn Mem Cogn* 39:297–303
- Kjelstrup KG, Tuvnes FA, Steffenach H-A, Murison R, Moser EI, Moser M-B (2002) Reduced fear expression after lesions of the ventral hippocampus. *Proc Natl Acad Sci USA* 99:10825–10830

- Kjelstrup KB, Solstad T, Brun VH, Hafting T, Leutgeb S, Witter MP, Moser EI, Moser M-B (2008) Finite scale of spatial representation in the hippocampus. *Science* 321:140–143
- Knierim JJ, Kudrimoti HS, McNaughton BL (1995) Place cells, head direction cells, and the learning of landmark stability. *J Neurosci* 15:1648–1659
- Kocsis B, Bragin A, Buzsáki G (1999) Interdependence of multiple theta generators in the hippocampus: a partial coherence analysis. *J Neurosci* 19:6200–6212
- Komorowski RW, Garcia CG, Wilson A, Hattori S, Howard MW, Eichenbaum H (2013) Ventral hippocampal neurons are shaped by experience to represent behaviorally relevant contexts. *J Neurosci* 33:8079–8087
- Konopacki J, Bland BH, Roth SH (1988) Carbachol-induced EEG “theta” in hippocampal formation slices: evidence for a third generator of theta in CA3c area. *Brain Res* 451:33–42
- Kosslyn S, Pick HL Jr, Fariello G (1974) Cognitive maps in children and men. *Child Dev* 45:707–716
- Kraus BJ, Robinson RJ, White JA, Eichenbaum H, Hasselmo ME (2013) Hippocampal “time cells”: time versus path integration. *Neuron* 78:1090–1101
- Krupic J, Bauza M, Burton S, Barry C, O’Keefe J (2015) Grid cell symmetry is shaped by environmental geometry. *Nature* 518:232–235
- Kurby CA, Zacks JM (2008) Segmentation in the perception and memory of events. *Trends Cogn Sci* 12:72–79
- Lánský P, Fenton AA, Vaillant J (2001) The overdispersion in activity of place cells. *Neurocomputing* 38–40:1393–1399
- Lee MG, Chrobak JJ, Sik A, Wiley RG, Buzsáki G (1994) Hippocampal theta activity following selective lesion of the septal cholinergic system. *Neuroscience* 62:1033–1047
- Lee H, Ghim J-W, Kim H, Lee D, Jung M (2012) Hippocampal neural correlates for values of experienced events. *J Neurosci* 32:15053–15065
- Lengyel M, Szatmáry Z, Erdi P (2003) Dynamically detuned oscillations account for the coupled rate and temporal code of place cell firing. *Hippocampus* 13:700–714
- Lengyel M, Kwag J, Paulsen O, Dayan P (2005) Matching storage and recall: hippocampal spike timing-dependent plasticity and phase response curves. *Nat Neurosci* 8:1677–1683
- Leung LS, Yu HW (1998) Theta-frequency resonance in hippocampal CA1 neurons in vitro demonstrated by sinusoidal current injection. *J Neurophysiol* 79:1592–1596
- Leutgeb S, Leutgeb JK, Treves A, Moser M-B, Moser EI (2004) Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science* 305:1295–1298
- Lever C, Burton S, Jeewajee A, O’Keefe J, Burgess N (2009) Boundary vector cells in the subiculum of the hippocampal formation. *J Neurosci* 29:9771–9777
- Lewis PR, Shute CC (1967) The cholinergic limbic system: projections to hippocampal formation, medial cortex, nuclei of the ascending cholinergic reticular system, and the subfornical organ and supra-optic crest. *Brain* 90:521–540
- Li XG, Somogyi P, Ylinen A, Buzsáki G (1994) The hippocampal CA3 network: an in vivo intracellular labeling study. *J Comp Neurol* 339:181–208
- Lorente De Nó R (1934) Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. *J Psychol Neurol* 46:113–117
- Losonczy A, Zemelman BV, Vaziri A, Magee JC (2010) Network mechanisms of theta related neuronal activity in hippocampal CA1 pyramidal neurons. *Nat Neurosci* 13:967–972
- Lubenov EV, Siapas AG (2009) Hippocampal theta oscillations are travelling waves. *Nature* 459:534–539
- Magee JC (2001) Dendritic mechanisms of phase precession in hippocampal CA1 pyramidal neurons. *J Neurophysiol* 86:528–532
- Malhotra S, Cross RWA, van der Meer MAA (2012) Theta phase precession beyond the hippocampus. *Rev Neurosci* 23:39–65
- Markus E, Qin Y, Leonard B, Skaggs W, McNaughton B, Barnes C (1995) Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *J Neurosci* 15:7079–7094

- Maurer AP, Vanrhoads SR, Sutherland GR, Lipa P, McNaughton BL (2005) Self-motion and the origin of differential spatial scaling along the septo-temporal axis of the hippocampus. *Hippocampus* 15:841–852
- Maurer AP, Cowen SL, Burke SN, Barnes CA, McNaughton BL (2006) Phase precession in hippocampal interneurons showing strong functional coupling to individual pyramidal cells. *J Neurosci* 26:13485–13492
- Maurer AP, Burke SN, Lipa P, Skaggs WE, Barnes CA (2012) Greater running speeds result in altered hippocampal phase sequence dynamics. *Hippocampus* 22:737–747
- McKenzie S, Robinson NTM, Herrera L, Churchill JC, Eichenbaum H (2013) Learning causes reorganization of neuronal firing patterns to represent related experiences within a hippocampal schema. *J Neurosci* 33:10243–10256
- McKenzie S, Frank AJ, Kinsky NR, Porter B, Rivière PD, Eichenbaum H (2014) Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron* 83:202–215
- McNamara T (1986) Mental representations of spatial relations. *Cogn Psychol* 18:87–121
- Mehta MR, Barnes CA, McNaughton BL (1997) Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc Natl Acad Sci USA* 94:8918–8921
- Mensink G, Raaijmakers J (1988) A model for interference and forgetting. *Psychol Rev* 95:434–455
- Mitchell SJ, Ranck JB (1980) Generation of theta rhythm in medial entorhinal cortex of freely moving rats. *Brain Res* 189:49–66
- Montello D (1991) The measurement of cognitive distance: methods and construct validity. *J Environ Psychol* 11:101–122
- Moser MB, Moser EI, Forrest E, Andersen P, Morris RG (1995) Spatial learning with a minislab in the dorsal hippocampus. *Proc Natl Acad Sci USA* 92:9697–9701
- Muller RU, Kubie JL (1987) The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J Neurosci* 7:1951–1968
- Muller RU, Kubie JL, Ranck JB (1987) Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *J Neurosci* 7:1935–1950
- O’Keefe J, Burgess N (1996) Geometric determinants of the place fields of hippocampal neurons. *Nature* 381:425–428
- O’Keefe J, Burgess N (2005) Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus* 15:853–866
- O’Keefe J, Nadel L (1978) *The hippocampus as a cognitive map*. Clarendon, New York, NY
- O’Keefe J, Recce ML (1993) Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3:317–330
- Pastalkova E, Itskov V, Amarasingham A, Buzsáki G (2008) Internally generated cell assembly sequences in the rat hippocampus. *Science* 321:1322–1327
- Patel J, Fujisawa S, Berényi A, Royer S, Buzsáki G (2012) Traveling theta waves along the entire septotemporal axis of the hippocampus. *Neuron* 75:410–417
- Pentkowski NS, Blanchard DC, Lever C, Litvin Y, Blanchard RJ (2006) Effects of lesions to the dorsal and ventral hippocampus on defensive behaviors in rats. *Eur J Neurosci* 23:2185–2196
- Petsche H, Stumpf C, Gogolak G (1962) The significance of the rabbit’s septum as a relay station between the midbrain and the hippocampus. I. The control of hippocampus arousal activity by the septum cells. *Electroencephalogr Clin Neurophysiol* 14:202–211
- Peyrache A, Lacroix MM, Petersen PC, Buzsáki G (2015) Internally organized mechanisms of the head direction sense. *Nat Neurosci* 18:569–575
- Pezzulo G, van der Meer MAA, Lansink CS, Pennartz CMA (2014) Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn Sci* 18:647–657
- Ranck J (1984) Head direction cells in the deep cell layer of dorsal presubiculum in freely moving rats. *Soc Neurosci Abstr* 10:599
- Ravassard P, Kees A, Willers B, Ho D, Aharoni D, Cushman J, Aghajan ZM, Mehta MR (2013) Multisensory control of hippocampal spatiotemporal selectivity. *Science* 340:1342–1346

- Redish AD, McNaughton BL, Barnes CA (2000a) Place cell firing shows an inertia-like process. *Neurocomputing* 32–33:235–241
- Redish AD, Rosenzweig ES, Bohanick JD, McNaughton BL, Barnes CA (2000b) Dynamics of hippocampal ensemble activity realignment: time versus space. *J Neurosci* 20:9298–9309
- Rich PD, Liaw H-P, Lee AK (2014) Place cells. Large environments reveal the statistical structure governing hippocampal representations. *Science* 345:814–817
- Rivard B, Li Y, Lenck-Santini P-P, Poucet B, Muller RU (2004) Representation of objects in space by two classes of hippocampal pyramidal cells. *J Gen Physiol* 124:9–25
- Robbe D, Buzsáki G (2009) Alteration of theta timescale dynamics of hippocampal place cells by a cannabinoid is associated with memory impairment. *J Neurosci* 29:12597–12605
- Robbe D, Montgomery SM, Thome A, Rueda-Orozco PE, McNaughton BL, Buzsáki G (2006) Cannabinoids reveal importance of spike timing coordination in hippocampal function. *Nat Neurosci* 9:1526–1533
- Rolls ET (2013) A quantitative theory of the functions of the hippocampal CA3 network in memory. *Front Cell Neurosci* 7:98
- Royer S, Sirota A, Patel J, Buzsáki G (2010) Distinct representations and theta dynamics in dorsal and ventral hippocampus. *J Neurosci* 30:1777–1787
- Royer S, Zemelman BV, Losonczy A, Kim J, Chance F, Magee JC, Buzsáki G (2012) Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. *Nat Neurosci* 15:769–775
- Rudell AP, Fox SE, Ranck JB (1980) Hippocampal excitability phase-locked to the theta rhythm in walking rats. *Exp Neurol* 68:87–96
- Sargolini F, Fyhn M, Hafting T, McNaughton BL, Witter MP, Moser M-B, Moser EI (2006) Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* 312:758–762
- Schmidt B, Papale A, Redish AD, Markus EJ (2013) Conflict between place and response navigation strategies: effects on vicarious trial and error (VTE) behaviors. *Learn Mem* 20:130–138
- Shankar KH, Howard MW (2015) Neural mechanism to simulate a scale-invariant future timeline. *arXiv preprint:1503.03322*
- Sik A, Penttonen M, Ylino A, Buzsáki G (1995) Hippocampal CA1 interneurons: an in vivo intracellular labeling study. *J Neurosci* 15:6651–6665
- Skaggs WE, McNaughton BL, Wilson MA, Barnes CA (1996) Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6:149–172
- Sloviter RS, Lømo T (2012) Updating the lamellar hypothesis of hippocampal organization. *Front Neural Circ* 6:102
- Solstad T, Boccara CN, Kropff E, Moser M-B, Moser EI (2008) Representation of geometric borders in the entorhinal cortex. *Science* 322:1865–1868
- Sreekumar V, Dennis S, Doxas I, Zhuang Y, Belkin M (2014) The geometry and dynamics of lifelogs: discovering the organizational principles of human experience. *PLoS One* 9:e97166, Balasubramaniam R (ed)
- Stark E, Eichler R, Roux L, Fujisawa S, Rotstein HG, Buzsáki G (2013) Inhibition-induced theta resonance in cortical circuits. *Neuron* 80:1263–1276
- Stensola T, Stensola H, Moser M-B, Moser EI (2015) Shearing-induced asymmetry in entorhinal grid cells. *Nature* 518:207–212
- Strange BA, Witter MP, Lein ES, Moser EI (2014) Functional organization of the hippocampal longitudinal axis. *Nat Rev Neurosci* 15:655–669
- Tamamaki N, Nojyo Y (1990) Disposition of the slab-like modules formed by axon branches originating from single CA1 pyramidal neurons in the rat hippocampus. *J Comp Neurol* 291:509–519
- Tamamaki N, Nojyo Y (1991) Crossing fiber arrays in the rat hippocampus as demonstrated by three-dimensional reconstruction. *J Comp Neurol* 303:435–442

- Taube JS, Burton HL (1995) Head direction cell activity monitored in a novel environment and during a cue conflict situation. *J Neurophysiol* 74:1953–1971
- Taube JS, Muller RU, Ranck JB (1990) Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J Neurosci* 10:420–435
- Tchumatchenko T, Clopath C (2014) Oscillations emerging from noise-driven steady state in networks with electrical synapses and subthreshold resonance. *Nat Commun* 5:5512
- Thurley K, Hellmundt F, Leibold C (2013) Phase precession of grid cells in a network model without external pacemaker. *Hippocampus* 23:786–796
- Traub RD, Miles R, Wong RK (1989) Model of the origin of rhythmic population oscillations in the hippocampal slice. *Science* 243:1319–1325
- Tsodyks MV, Skaggs WE, Sejnowski TJ, McNaughton BL (1996) Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus* 6:271–280
- Tulving E, Markowitsch HJ (1998) Episodic and declarative memory: role of the hippocampus. *Hippocampus* 8:198–204
- Unsworth N (2008) Exploring the retrieval dynamics of delayed and final free recall: further evidence for temporal-contextual search. *J Mem Lang* 59:223–236
- Vaidya SP, Johnston D (2013) Temporal synchrony and gamma-to-theta power conversion in the dendrites of CA1 pyramidal neurons. *Nat Neurosci* 16:1812–1820
- Van der Meer MAA, Redish AD (2011) Theta phase precession in rat ventral striatum links place and reward information. *J Neurosci* 31:2843–2854
- Vanderwolf CH (1969) Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalogr Clin Neurophysiol* 26:407–418
- Vazdarjanova A, Guzowski JF (2004) Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. *J Neurosci* 24:6489–6496
- Wang Y, Romani S, Lustig B, Leonardo A, Pastalkova E (2014) Theta sequences are essential for internally generated hippocampal firing fields. *Nat Neurosci* 18:282–288
- White JA, Banks MI, Pearce RA, Kopell NJ (2000) Networks of interneurons with fast and slow gamma-aminobutyric acid type A (GABAA) kinetics provide substrate for mixed gamma-theta rhythm. *Proc Natl Acad Sci USA* 97:8128–8133
- Whitlock JR, Derdikman D (2012) Head direction maps remain stable despite grid map fragmentation. *Front Neural Circ* 6:9
- Wikenheiser AM, Redish AD (2015) Hippocampal theta sequences reflect current goals. *Nat Neurosci* 18:289–294
- Wikenheiser AM, Stephens DW, Redish AD (2013) Subjective costs drive overly patient foraging strategies in rats on an intertemporal foraging task. *Proc Natl Acad Sci USA* 110:8308–8313
- Wittner L, Henze DA, Záborszky L, Buzsáki G (2007) Three-dimensional reconstruction of the axon arbor of a CA3 pyramidal cell recorded and filled in vivo. *Brain Struct Funct* 212:75–83
- Wu X, Foster DJ (2014) Hippocampal replay captures the unique topological structure of a novel environment. *J Neurosci* 34:6459–6469
- Wu MV, Hen R (2014) Functional dissociation of adult-born neurons along the dorsoventral axis of the dentate gyrus. *Hippocampus* 24:751–761
- Zhang K, Ginzburg I, McNaughton BL, Sejnowski TJ (1998) Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J Neurophysiol* 79:1017–1044
- Zugaro MB, Monconduit L, Buzsáki G (2005) Spike phase precession persists after transient intrahippocampal perturbation. *Nat Neurosci* 8:67–71

Cortical Evolution: Introduction to the Reptilian Cortex

Gilles Laurent, Julien Fournier, Mike Hemberger, Christian Müller, Robert Naumann, Janie M. Ondracek, Lorenz Pammer, Samuel Reiter, Mark Shein-Idelson, Maria Antonietta Tosches, and Tracy Yamawaki

Abstract Some 320 million years ago (MYA), the evolution of a protective membrane surrounding the embryo, the amnion, enabled vertebrates to develop outside water and thus invade new terrestrial niches. These amniotes were the ancestors of today's mammals and sauropsids (reptiles and birds). Present-day reptiles are a diverse group of more than 10,000 species that comprise the sphenodon, lizards, snakes, turtles and crocodylians. Although turtles were once thought to be the most “primitive” among the reptiles, current genomic data point toward two major groupings: the Squamata (lizards and snakes) and a group comprising both the turtles and the Archosauria (dinosaurs and modern birds and crocodiles). Dinosaurs inhabited the Earth from the Triassic (230 MYA), at a time when the entire landmass formed a single Pangaea. Dinosaurs flourished from the beginning of the Jurassic to the mass extinction at the end of the Cretaceous (65 MYA), and birds are their only survivors. What people generally call reptiles is thus a group defined in part by exclusion: it gathers amniote species that are neither mammals nor birds, making the reptiles technically a paraphyletic grouping. Despite this, the so-defined reptiles share many evolutionary, anatomical, developmental, physiological (e.g., ectothermia), and functional features. It is thus reasonable to talk about a “reptilian brain.”

Reptilian Brain Structure and Evolution

The diversity of reptiles and their evolutionary relationship to mammals make reptilian brains great models to explore questions related to the structural and functional evolution of vertebrate neural circuits. To this end, comparative studies

G. Laurent (✉) • J. Fournier • M. Hemberger • C. Müller • R. Naumann • J.M. Ondracek • L. Pammer • S. Reiter • M. Shein-Idelson • M.A. Tosches • T. Yamawaki
Max Planck Institute for Brain Research, Max von Laue Str. 4, 60438 Frankfurt am Main, Germany
e-mail: gilles.laurent@brain.mpg.de

seek to identify homologies, i.e., structural or molecular similarities due to common ancestry, at a variety of levels, e.g., brain regions, circuits, or cell types. Homologies can be inferred from extant species by using a comparative approach within a phylogenetic framework. Vertebrate brains have been classically compared in terms of morphology, connectivity, and neurochemistry. However, adult neuroanatomy may not be sufficient to determine homologies without ambiguity.

Vertebrates share among themselves the highest morphological resemblance not as adults but at their “phylogenic stage” of embryonic development, as mirrored by similarity across transcriptomes. Identification of conserved brain subdivisions, established by conserved signaling centers and uniquely defined by the combinatorial expression of transcription factors during development, demonstrates that all of the general brain regions found in mammals, including the cerebral cortex, have homologies in reptiles. For example, expression of transcription factors such as *Emx1* and *Tbr1* in developing frogs, turtles, chickens, and mice reveals the existence of the same fundamental subdivisions of the pallium (ventral, medial, dorsal and lateral) despite the divergent morphologies of pallial structures in adults. To be clear, the cerebral cortex is thus not a mammalian invention but rather an ancient dorsal pallial structure that predates the split between sauropsids and therapsids (the mammals’ precursors) (Fournier et al. 2015; Naumann et al. 2015).

Comparative studies of brain structure and development have revealed a general bauplan that describes the fundamental large-scale architecture of the vertebrate brain and provides insight into its basic functional organization. The telencephalon not only integrates and stores multimodal information but is also the higher center of action selection and motor control (basal ganglia). The hypothalamus is a conserved area controlling homeostasis and behaviors essential for survival, such as feeding and reproduction. Furthermore, in all vertebrates, behavioral states are controlled by common brainstem neuromodulatory circuits, such as the serotonergic system. Finally, vertebrates harbor a diverse set of sense organs, and their brains share pathways for processing incoming sensory inputs. For example, in all vertebrates, visual information from the retina is relayed and processed to the pallium through the tectum and the thalamus, whereas olfactory input from the nose first reaches the olfactory bulb (OB) and then the pallium.

Although pallial structures exist in amphibians and fish, reptiles and mammals are the only vertebrates to possess a cerebral cortex with a clear, though simple, three-layered structure similar to that of mammalian allocortex. The reptilian ventral pallium also gives rise to the dorsal ventricular ridge (DVR), a structure that dominates the bird pallium and contributes to the complex cognitive abilities of birds, but whose mammalian equivalent is still the subject of debate among comparative anatomists. The reptilian cortex contains far fewer subdivisions than that of rodents, carnivores, or primates: it is subdivided into a medial cortex, often called hippocampus by anatomists, a lateral cortex, equivalent to the mammalian piriform cortex, and a dorsal cortex in between, which receives multimodal inputs (e.g., visual in turtles). There is little evidence for motor and somatosensory areas in the reptilian cortex, but pallial motor control may have evolved early in vertebrate evolution. Owing to this simplicity, the reptilian brain facilitates the study of

primordial cortical function as a whole and points to the origins of cortex as fulfilling general associative functions.

Besides sharing pallial modules, mammals and reptiles also share a complement of cortical cell types, suggesting that some structural elements of cortical circuits arose early in amniote evolution. As in mammals, reptilian cortex contains excitatory, glutamatergic neurons and inhibitory, GABAergic interneurons. In both mammals and reptiles, these neurons have a common developmental origin: excitatory neurons are generated by multipotent cortical progenitors, whereas inhibitory neurons are born in the subpallium before migrating to the cortex.

Classical studies suggest that the reptilian main cortical cell layer (layer 2) corresponds to the deep, output layers of mammalian neocortex, whereas its layer 1 is equivalent to mammalian layer I. According to this view, mammalian cortical evolution would have included the incorporation of new, intermediate cell layers acting as input stations and internal circuitry. Morphologically, L2 pyramidal neurons of the reptilian dorsal cortex are most similar to mammalian hippocampal excitatory neurons. Indeed, reptilian pyramidal neurons have, depending on the area, little to no basal dendritic field and several densely spine-studded apical dendrites, quite different from the single, long, apical dendrite of neocortical pyramidal neurons. Consistent with this correspondence between layers, reptilian subpallial cells transplanted into mammalian embryos generate GABAergic neurons that can populate only the deeper cortical layers.

Challenging this view, however, recent molecular studies show that turtle and lizard cortical neuroblasts generate neurons that express upper layer molecular markers in a developmental sequence similar to that observed in mammals. Although the molecular characterization of neuronal types in the reptilian cortex is still in its infancy, it is possible that reptilian cortex represents an ancestral blueprint for the more elaborate mammalian cortical circuits. For example, reptilian cortical neurons, or subsets of them, might share molecular (and functional) features with both upper- and lower-layer mammalian cells.

Basic Architecture and Functional Features

Vertical Connectivity

The architecture of PCx and DCx is archetypal of a three-layered paleocortex. Layer 1 contains mainly dendrites of layer 2 principal cells, a few scattered interneurons and afferent and local axons. Layer 2 contains the densely packed somata of pyramidal cells, whose apical dendrites run radially towards the pial surface. Layer 3 contains basal dendrites of pyramidal cells, corticofugal and local axons, some interneurons and a few deep pyramidal neurons in PCx (Neville and Haberly 2004; Ulinski 1990). Incoming afferents to PCx run through the lateral olfactory tract (LOT; Sosulski et al. 2011); those to DCx run through the lateral

forebrain bundle (LFB; Mulligan and Ulinski 1990). These input fibers fan out below the pial surface and make en-passant synapses on cortical neurons within the distal 50–100 μm of layer 1 (Haberly and Behan 1983; Smith et al. 1980). Afferent synapses impinge on both layer-1 interneurons and on distal dendrites of layer-2 pyramidal cells; interneurons provide both feed-forward and feedback inhibition to pyramidal cells that themselves provide recurrent excitation to other pyramidal neurons (Smith et al. 1980; Suzuki and Bekkers 2011, 2012; Kriegstein and Connors 1986; Mancilla et al. 1998). In both PCx and DCx, superficial layer-1 interneurons tend to receive a higher density of afferent input than pyramidal cells do (Smith et al. 1980; Suzuki and Bekkers 2012; Stokes and Isaacson 2010), which, combined with a strong feed-back inhibition via layer-2/3 interneurons (Suzuki and Bekkers 2012; Kriegstein and Connors 1986; Stokes and Isaacson 2010) may explain the observed strong inhibition evoked by sensory stimulation and the sparseness of pyramidal cell firing. To a first degree, PCx and DCx thus have a similar microcircuit layout: both exhibit distal dendritic excitation from sensory afferents, strong feed-forward inhibition, recurrent excitation through the so-called associational intracortical connections, and feedback inhibition (Haberly 2001; Shepherd 2011).

Different cell types have been identified in PCx. Most segregate into specific sub-layers of the piriform microcircuit. Excitatory neurons in layer 2 can be subdivided in semilunar (upper layer 2) and superficial pyramidal neurons (lower layer 2), whereas those in layer 3 comprise a few deep pyramidal cells and scattered multipolar spiny glutamatergic neurons (Haberly 1983; Suzuki and Bekkers 2006; Bekkers and Suzuki 2013). Although they are embedded in the same basic connectivity scheme, semilunar and superficial pyramidal cells receive different ratios of afferent to associational inputs and may therefore belong to distinct functional sub-circuits (Suzuki and Bekkers 2011; but see Poo and Isaacson 2011), consistent with morphological differences between their dendritic trees and their laminar position (Wiegand et al. 2011). Although data on subpopulations of principal cells in DCx are few, analysis of Golgi-stained material also revealed different morphological classes of spiny neurons at different laminar and sublaminar positions in reptilian cortex (Ulinski 1977; Desan 1984). PCx and DCx pyramidal neurons are also similar with respect to their dendritic electrophysiological properties, suggesting comparable integrative properties at the subcellular level (Larkum et al. 2008; Bathellier et al. 2009). Different subtypes of inhibitory interneurons have been identified in PCx based on molecular markers, the morphology of their dendritic arbor and the distribution of their axonal projections (reviewed in Suzuki and Bekkers 2007). These sub-classes seem to correlate with the type of inhibition they subserve, i.e., primarily feedback or feed-forward. Horizontal and neurogliaform interneurons in layer 1 receive afferent inputs from the LOT and mediate fast feed-forward inhibition targeting apical dendrites of layer-2 pyramidal cells. Bitufted, fast-spiking and regular spiking interneurons from layers 2 and 3 receive very little direct afferent input from the LOT but provide strong feedback inhibition onto the somata and basal dendrites of pyramidal cells (Suzuki and Bekkers 2012; Stokes and Isaacson 2010). Similarly, different

populations of inhibitory interneurons in turtle DCx subserve mainly feed-forward (subpial cells; Mancilla et al. 1998) or feedback (Mancilla et al. 1998; Connors and Kriegstein 1986) inhibition. Axonal reconstructions of DCx interneurons (Colombe et al. 2004) and immunocytochemical labeling (Reiner 1991, 1993) suggest the existence of morphologically and physiologically identifiable classes of inhibitory interneurons. It remains to be shown that those groupings also share functional similarities with those in PCx. Given the anatomical similarity of input projections to PCx and DCx, one may speculate that the inhibitory circuit topology of these two cortices could also be similar.

Horizontal Connectivity

In PCx, afferents from mitral/tufted (MT) cells appear to project throughout the cortex without any clear topographical relationship to their glomeruli of origin (Sosulski et al. 2011; Miyamichi et al. 2011; Illig and Haberly 2003; Apicella et al. 2010; Ghosh et al. 2011). Although this does not rule out the possibility of some fine-scale topographical mapping of OB projections (e.g., mitral vs. tufted cell projections) (Igarashi et al. 2012), it is now accepted that the glomerular clustering of olfactory receptor cells axons in OB is entirely discarded at the level of PCx (Wilson and Sullivan 2011). In DCx, early tracing studies from Ulinski and colleagues suggested that the visual field is projected onto the rostro-caudal axis of DCx in the form of iso-azimuth lamellae covering the naso-temporal dimension of the visual field (Mulligan and Ulinski 1990; Ulinski and Nautiyal 1988). Such a mapping of projections still awaits physiological confirmation and fine thalamo-cortical projection tracing. If confirmed, this topographical mapping would differ from the topology of mammalian olfactory projections to PCx, at least along one cortical dimension.

In both PCx and DCx, the density of sensory afferents varies over the cortical surface: high rostrally and laterally, it decreases progressively as one moves away from the entry point of the LOT (PCx) or the LFB (DCx). Hence, the balance between afferent and associational connectivity decreases along the rostro-caudal and latero-medial (or ventro-dorsal) axes (Mulligan and Ulinski 1990; Haberly 2001; Wilson and Sullivan 2011; Hagiwara et al. 2012; Cosans and Ulinski 1990). PCx is subdivided into anterior and posterior regions, which differ not only in the density of afferent vs. associational fibers (Haberly 2001) but also in the properties of odor-evoked responses (Litaudon et al. 2003; Kadohisa and Wilson 2006). PCx microcircuits may also contain fine-grain connectivity gradients: in vitro recordings from aPCx reveal that inhibition of pyramidal cells is asymmetric and stronger along the rostro-caudal axis of the anterior part of PCx, over distances as short as 200 μm (Luna and Pettit 2010). In turtles, DCx has been classically divided into two different regions (D2 and D1) along the latero-medial axis (Ulinski 1990; Desan 1984). This dichotomy rests mostly on cytoarchitectural features related to the thickness of subcellular layer 3: thick in D2 laterally, thin in

D1, with a significant transition zone between the two. Recent molecular data suggest that this separation may be correlated with a higher expression level of layer-4 markers in D2 (Dugas-ford et al. 2012). Confirmation of this division and of its potential functional significance needs additional work. Such gradients of connectivity across the cortical surface (in PCx and DCx) should be clearly described because any horizontal heterogeneity could influence the propagation and reverberation of activity across cortex, under the combined influences of spreading afferent input and widespread associational activity.

Given their reciprocal interconnections with high-order cortical areas and a lack of evident sensory topography, PCx and DCx are sometime described as associational rather than primary sensory cortices (Shepherd 2011). The major partners of PCx are the orbitofrontal cortex (Ekstrand et al. 2001; Illig 2006), the lateral entorhinal cortex (Kerr et al. 2007; Johnson et al. 2000) and the agranular insular cortex (Johnson et al. 2000). Connectivity to these downstream targets differs between aPCx and pPCx, supporting the notion that they have different functions. Similarly, DCx is reciprocally connected to dorso-medial (DMCx) and medial (MCx) cortices (Ulinski 1977; Desan 1984). Those regions are, on the basis of hodology and position, often compared to parahippocampal and hippocampal cortices (Desan 1984; Northcutt 1981; López et al. 2003; Aboitiz et al. 2003). Both PCx and DCx are thus directly connected to associational networks likely involved in controlling or modulating behavior.

PCx and DCx are further interconnected with other cortical-like areas that also receive parallel sensory afferents from the OB or the lateral geniculate nucleus of the thalamus (LGN), respectively. For PCx, these include the anterior olfactory nucleus (AON; Haberly and Price 1978; Illig and Eudy 2009), the olfactory tubercle (OT; Haberly and Price 1978), and the amygdala (Johnson et al. 2000; Luna and Morozov 2012). AON might be a first stage of odorant-feature processing, in turn used by PCx to detect complex odorant combinations (Haberly 2001; Lei et al. 2006; Kay et al. 2011). DCx's AON equivalent could be the pallial thickening (PT), for it receives direct thalamic afferent input and projects to DCx (Mulligan and Ulinski 1990; Heller and Ulinski 1987). If AON and PT also share functional characteristics, these similarities may point to common elementary processing streams of three-layered sensory cortices.

In turtles, visual stimulation triggers propagating waves of neural activity that travel across the cortex. These waves are slower and simpler than those observed in mammalian neocortex. They are accompanied by relatively slow oscillations, which are most prominent in the 20 Hz frequency band. Whereas the so-called gamma oscillations in mammalian cortex are typically around and above 40 Hz, recent results in mice indicate that the 20 Hz band dominates when parvalbumin (PV) interneuron development is artificially arrested, consistent with the above observation that turtle cortex lacks PV interneurons. The computational role, if any, of such dynamics is unknown at present. Progress will require new experimental approaches that allow the simultaneous sampling of large neuronal populations. Specific and data-driven theories of computation in reptilian cortex thus await further study. To the extent that modern reptilian cortex resembles that in the

common ancestor of reptiles and mammals, understanding reptilian cortex function may reveal some of the fundamental associative computations that early cortical circuits evolved to carry out.

Physiological Adaptations of the Reptilian Brain

Reptiles are ectotherms and must behaviorally or physiologically adapt to varying temperatures, making reptile brains notable for their adaptation to extreme conditions. The turtle brain, for example, has evolved remarkable adaptations to hypoxic conditions, which have long been known to experimentalists interested in hearing, olfaction, motor control or cerebellar physiology. Semi-aquatic freshwater turtles (e.g., *Chrysemys picta*) are remarkable for their ability to survive hibernation in hypoxic waters during cold northern winters and have been reported to survive anoxic periods of over 4 months at 3 °C. This resilience points to several physiological adaptations to survive cold, anoxic conditions during hibernation and subsequent re-oxygenation in the spring. The mechanisms underlying such tolerance are interesting—and also possibly applicable to the treatment of brain injury due to ischemia.

Anoxia tolerance critically depends on the ability to reduce energy expenditure while elevating anaerobic adenosine triphosphate (ATP) production. ATP consumption in neural tissues, particularly those involved in sensory functions, requires high ATP turnover. Free glucose and glycogen can serve as metabolic substrates in anoxic conditions, and stores of glycogen are released from the liver and skeletal muscles to facilitate ATP production during hibernation. But by shifting to anaerobic glycolysis, hibernating turtles must buffer the resulting metabolic product, lactic acid, to prevent fatal acidosis. In turtles, extracellular buffering of lactic acid is achieved by the large volumes of pericardial and peritoneal fluids and through sequestration in the mineralized shell and skeleton.

During hibernation, turtles reduce ATP consumption in neural tissue by reducing membrane permeability and, thus, the need for pump-mediated restoration of ionic balance. This phenomenon has been observed in turtle neurons for voltage-gated Na^+ channels, the oxygen-sensitive calcium-activated K^+ channels, as well as AMPA and NMDA receptors. Neuronal depolarization and NMDA receptor over-activation are hallmarks of anoxia in mammalian neural tissue and lead to excitotoxic cell death via increased intracellular Ca^{2+} concentrations. In turtle neurons, NMDA receptor opening times are reduced by 65 % after 60 min of anoxia, and NMDA receptor currents are actually silenced by anoxia-triggered cytosolic Ca^{2+} release from mitochondrial stores. Despite reduced channel activity and, thus, reduced electrical activity during anoxia, turtles appear to remain vigilant and responsive to visual stimuli during hibernation (but less so to tactile stimuli). This maintenance of visual responses in the absence of tactile responses during hibernation suggests a differential down-regulation of sensory modalities. Because changes in daily illumination contribute to signaling the end of hibernation, it may

be a selective advantage for energy-conserving mechanisms to be applied less to the visual system than to other brain areas.

When oxygen again becomes available, the re-oxygenation of tissue, such as that observed after ischemia-reperfusion events caused by myocardial infarction or stroke, is problematic in mammals because it is associated with the excessive production of reactive oxygen species (ROS) that damage cells through protein oxidation and lipid peroxidation. Turtles have evolved effective physiological mechanisms that allow for repeated cycles of anoxia and re-oxygenation without the free-radical damage induced by ischemia-reperfusion of tissues observed in mammals. These mechanisms involve an evolutionarily old up-regulation of nitric oxide and its metabolites that limits ROS generation and tissue damage in cardiac and brain tissue. Thus, the turtle brain, with these physiological adaptations, is a very useful model for the study of neural circuit function. Neural tissue remains alive and viable for days *ex vivo*, and tissue can be routinely stored overnight in a refrigerator at 3–5 °C.

Brain and Behavior

Reptiles express a number of complex behaviors normally attributed to mammals. They can, for example, learn to navigate mazes as well as birds or mammals and likely use a hippocampal structure to do so. Remarkably, sea turtles navigate across the ocean to their home beach, possibly using magnetic, olfactory, and visual cues. Reptiles also express interesting social behaviors: pythons and crocodiles, for example, incubate their eggs after hatching, and mothers guard the nest and provide infant care. Some lizard species exhibit pair bonding and parental care and form families. Others construct networks of tunnels and may live in social communities like those of naked-mole rats.

Reptiles inhabit a variety of ecological niches, such as terrestrial deserts, temperate forests, tropical jungles, the sea, and fresh waters. They are found in all oceans and on all continents except Antarctica. Semi-aquatic turtles, which live in primarily aquatic habitats but make terrestrial excursions to lay eggs, offer an interesting model animal in which to study navigation. Many turtles are known for their impressive navigational skills and high nest-site fidelity, or the tendency for individuals to return to the same geographic location or microhabitat to nest. The mechanisms underlying these behaviors are not entirely understood, but navigational ability in the fresh water turtle *Chrysemys picta* seems to depend on experience gained during a critical period (<4 years of age). Translocated resident adults, for example, can quickly and consistently find new aquatic habitats within 33 h, often using precise and predictable routes, whereas translocated non-resident adults appear unable to find new aquatic habitats even after 21 days.

The hippocampus is critical for map-like or relational memory representations of allocentric space, where objects are represented in relation to other objects. In birds and mammals, lesions to the hippocampal formation produce selective impairments

in spatial tasks that require the encoding of relationships among multiple environmental features (place learning) but not in tasks that require approaching a single cue or simple non-spatial discriminations. Whereas extensive comparative research supports the idea that the reptilian medial cortex is homologous to the hippocampal formation of mammals and birds, only a few studies have examined the neural function of this brain structure or its role in place learning. In one such study, Rodríguez et al. (2002) evaluated the effects of lesions to the hippocampus of turtles in place and cue-maze tasks. Hippocampus-lesioned (and sham-lesioned) animals performed cue-discrimination tasks correctly but hippocampus-lesioned animals failed at the place learning that relied on allocentric space learning. These results indicate that lesions to the hippocampus of turtles selectively impair map-like memory representations of the environmental space, mirroring the effect of hippocampal lesions in mammals and birds. Thus reptilian hippocampus may also share a central role in navigation.

In conclusion, the observation that mammalian and reptilian brains share both ancestry and a large number of functional attributes suggests that the identification of primordial (and possibly general) algorithmic principles of brain function could be helped by comparative approaches. To this end, the reptilian brain, with its simpler structure, may prove invaluable to decipher fundamental questions of modern neuroscience.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Aboitiz F, Morales D, Montiel J (2003) The evolutionary origin of the mammalian isocortex: towards an integrated developmental and functional approach. *Behav Brain Sci* 26:535–552
- Apicella A, Yuan Q, Scanziani M, Isaacson JS (2010) Pyramidal cells in piriform cortex receive convergent input from distinct olfactory bulb glomeruli. *J Neurosci* 30:14255–14260
- Bathellier B, Margrie TW, Larkum ME (2009) Properties of piriform cortex pyramidal cell dendrites: implications for olfactory circuit design. *J Neurosci* 29:12641–12652
- Bekkers JM, Suzuki N (2013) Neurons and circuits for odor processing in the piriform cortex. *Trends Neurosci* 36:429–438
- Colombe JB, Sylvester J, Block J (2004) Subpial and stellate cells: two populations of interneurons in turtle visual cortex. *J Comp Neurol* 471:333–351
- Connors BW, Kriegstein AFL (1986) Cellular physiology of the turtle visual cortex: distinctive properties of pyramidal and stellate neurons. *J Neurosci* 6:164–177

- Cosans CE, Ulinski PS (1990) Spatial organization of axons in turtle visual cortex: intralamellar and interlamellar projections. *J Comp Neurol* 296:548–558
- Desan PH (1984) The organization of the cerebral cortex of the pond turtle, *Pseudemys scripta elegans*. PhD thesis, Harvard University, Cambridge, MA
- Dugas-ford J, Rowell JJ, Ragsdale CW (2012) Cell-type homologies and the origins of the neocortex. *Proc Natl Acad Sci USA* 109:16974–16979
- Ekstrand JJ, Domroese ME, Johnson DMG, Feig SL, Knodel SM, Behan M, Haberly LB (2001) A new subdivision of anterior piriform cortex and associated deep nucleus with novel features of interest for olfaction and epilepsy. *J Comp Neurol* 434:289–307
- Fournier J, Müller CM, Laurent G (2015) Looking for the roots of cortical sensory computation in three-layered cortices. *Curr Opin Neurobiol* 31:119–126
- Ghosh S, Larson SD, Hefzi H, Marnoy Z, Cutforth T, Dokka K, Baldwin KK (2011) Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. *Nature* 472:217–220
- Haberly LB (1983) Structure of the piriform cortex of the opossum. I. Description of neuron types with Golgi methods. *J Comp Neurol* 213:163–187
- Haberly LB (2001) Parallel-distributed processing in olfactory cortex: new insights from morphological and physiological analysis of neuronal circuitry. *Chem Senses* 26:551–576
- Haberly L, Behan M (1983) Structure of the piriform cortex of the opossum. III. Ultrastructural characterization of synaptic terminals of association and olfactory bulb afferent fibers. *J Comp Neurol* 219:448–460
- Haberly LB, Price JL (1978) Association and commissural fiber systems of the olfactory cortex of the rat. *J Comp Neurol* 178:711–740
- Hagiwara A, Pal SK, Sato TF, Wienisch M, Murthy VN, Shepherd GM (2012) Optophysiological analysis of associational circuits in the olfactory cortex. *Front Neural Circ* 6:18
- Heller SB, Ulinski PS (1987) Morphology of geniculocortical axons in turtles of the genera *Pseudemys* and *Chrysemys*. *Anat Embryol* 175:505–515
- Igarashi KM, Ieki N, An M, Yamaguchi Y, Nagayama S, Kobayakawa K, Kobayakawa R, Tanifuji M, Sakano H, Chen WR et al (2012) Parallel mitral and tufted cell pathways route distinct odor information to different targets in the olfactory cortex. *J Neurosci* 32:7970–7985
- Illig KR (2006) Projections from orbitofrontal cortex to anterior piriform cortex in the rat suggest a role in olfactory information processing. *J Comp Neurol* 488:224–231
- Illig KR, Eudy JD (2009) Contralateral projections of the rat anterior olfactory nucleus. *J Comp Neurol* 512:115–123
- Illig KR, Haberly LB (2003) Odor-evoked activity is spatially distributed in piriform cortex. *J Comp Neurol* 457:361–373
- Johnson DMG, Illig KR, Behan M, Haberly LB (2000) New features of connectivity in piriform cortex visualized by intracellular injection of pyramidal cells suggest that “primary” olfactory cortex functions like “association” cortex in other sensory systems. *J Neurosci* 20:6974–6982
- Kadohisa M, Wilson DA (2006) Separate encoding of identity and similarity of complex familiar odors in piriform cortex. *Proc Natl Acad Sci USA* 103:15206–15211
- Kay RB, Meyer EA, Illig KR, Brunjes PC (2011) Spatial distribution of neural activity in the anterior olfactory nucleus evoked by odor and electrical stimulation. *J Comp Neurol* 519:277–289
- Kerr KM, Agster KL, Furtak SC, Burwell RD (2007) Functional neuroanatomy of the parahippocampal region: the lateral and medial entorhinal areas. *Hippocampus* 17:697–708
- Kriegstein R, Connors BW (1986) Cellular physiology of the turtle visual cortex: synaptic properties and intrinsic circuitry. *J Neurosci* 6:178–191
- Larkum ME, Watanabe S, Lasser-ross N, Rhodes P, Ross WN, Ledergerber D, Larkum ME (2008) Dendritic properties of turtle pyramidal neurons. *J Neurophysiol* 99:683–694
- Lei H, Mooney R, Katz LC (2006) Synaptic integration of olfactory information in mouse anterior olfactory nucleus. *J Neurosci* 26:12023–12032
- Litaudon P, Amat C, Bertrand B, Vigouroux M, Buonviso N (2003) Piriform cortex functional heterogeneity revealed by cellular responses to odours. *Eur J Neurosci* 17:2457–2461

- López JC, Vargas JP, Gómez Y, Salas C (2003) Spatial and non-spatial learning in turtles: the role of medial cortex. *Behav Brain Res* 143:109–120
- Luna VM, Morozov A (2012) Input-specific excitation of olfactory cortex microcircuits. *Front Neural Circ* 6:1–7
- Luna VM, Pettit DL (2010) Asymmetric rostro-caudal inhibition in the primary olfactory cortex. *Nat Neurosci* 13:533–535
- Mancilla JG, Fowler M, Ulinski PS (1998) Responses of regular spiking and fast spiking cells in turtle visual cortex to light flashes. *Vis Neurosci* 15:979–993
- Miyamichi K, Amat F, Moussavi F, Wang C, Wickersham I, Wall NR, Taniguchi H, Tasic B, Huang ZJ, He Z et al (2011) Cortical representations of olfactory input by trans-synaptic tracing. *Nature* 472:191–196
- Mulligan KA, Ulinski PS (1990) Organization of geniculocortical projections in turtles: isoazimuth lamellae in the visual cortex. *J Comp Neurol* 296:531–547
- Naumann R, Ondracek JM, Reiter S, Shein-Idelson M, Tosches MA, Yamawaki T, Laurent G (2015) Reptilian brain primer. *Curr Biol* 25(8):R317–R321
- Neville KR, Haberly LB (2004) Olfactory cortex. In: Shepherd GM (ed) *The synaptic organization of the brain*. Oxford University Press, New York, NY, pp 415–454
- Northcutt RG (1981) Evolution of the telencephalon in nonmammals. *Annu Rev Neurosci* 4:301–350
- Poo C, Isaacson JS (2011) A major role for intracortical circuits in the strength and tuning of odor-evoked excitation in olfactory cortex. *Neuron* 72:41–48
- Reiner A (1991) A comparison of the neurotransmitter-specific and neuropeptide-specific neuronal cell types present in turtle cortex to those present in mammalian isocortex: implications for the evolution of isocortex. *Brain Behav Evol* 38:53–91
- Reiner A (1993) Neurotransmitter organization and connections of turtle cortex: implications for the evolution of mammalian isocortex. *Comp Biochem Physiol* 104:735–748
- Rodríguez F, López JC, Vargas JP, Gómez Y, Broglio C, Salas C (2002) Conservation of spatial memory function in the pallial forebrain of reptiles and ray-finned fishes. *J Neurosci* 22(7):2894–2903
- Shepherd GM (2011) The microcircuit concept applied to cortical evolution: from three-layer to six-layer cortex. *Front Neuroanat* 5:1–15
- Smith LM, Ebner FF, Colonnier M (1980) The thalamocortical projection in *Pseudemys* turtles: a quantitative electron microscopic study. *J Comp Neurol* 461:445–461
- Sosulski DL, Bloom ML, Cutforth T, Axel R, Datta SR (2011) Distinct representations of olfactory information in different cortical centres. *Nature* 472:213–216
- Stokes CCA, Isaacson JS (2010) From dendrite to soma: dynamic routing of inhibition by complementary interneuron microcircuits in olfactory cortex. *Neuron* 67:452–465
- Suzuki N, Bekkers JM (2006) Neural coding by two classes of principal cells in the mouse piriform cortex. *J Neurosci* 26:11938–11947
- Suzuki N, Bekkers JM (2007) Inhibitory interneurons in the piriform cortex. *Clin Exp Pharmacol Physiol* 34:1064–1069
- Suzuki N, Bekkers JM (2011) Two layers of synaptic processing by principal neurons in piriform cortex. *J Neurosci* 31:2156–2166
- Suzuki N, Bekkers JM (2012) Microcircuits mediating feedforward and feedback synaptic inhibition in the piriform cortex. *J Neurosci* 32:919–931
- Ulinski PS (1977) Intrinsic organization of snake medial cortex: an electron microscopic and Golgi study. *J Morphol* 152:247–279
- Ulinski PS (1990) The cerebral cortex of reptiles. *Cereb Cort* 8A:139–216
- Ulinski PS, Nautiyal J (1988) Organization of retinogeniculate projections in turtles of the genera *Pseudemys* and *Chrysemys*. *J Comp Neurol* 276:92–112
- Wiegand HF, Beed P, Bendels MHK, Leibold C, Schmitz D, Jochenning FW (2011) Complementary sensory and associative microcircuitry in primary olfactory cortex. *J Neurosci* 31:12149–12158
- Wilson DA, Sullivan RM (2011) Cortical processing of odor objects. *Neuron* 72:506–519

Flow of Information Underlying a Tactile Decision in Mice

Nuo Li, Zengcai V. Guo, Tsai-Wen Chen, and Karel Svoboda

Abstract Motor planning allows us to conceive, plan, and initiate skilled motor behaviors. Motor planning involves activity distributed widely across the cortex. How this activity dynamically comes together to guide movement remains an unsolved problem. We study motor planning in mice performing a tactile decision behavior. Head-fixed mice discriminate object locations with their whiskers and report their choice by directional licking (“lick left”/“lick right”). A short-term memory component separates tactile “sensation” and “action” into distinct epochs. Using loss-of-function experiments, cell-type specific electrophysiology, and cellular imaging, we delineate when and how activity in specific brain areas and cell types drives motor planning in mice. Our results suggest that information flows serially from sensory to motor areas during motor planning. The motor cortex circuit maintains the motor plan during short-term memory and translates the motor plan into motor commands that drive the upcoming directional licking.

Introduction

With a half a second’s planning, we rapidly carry out complex sequences of movements. Motor planning refers to our ability to conceive, plan, and initiate a skilled motor behavior. During motor planning, sensory information must be integrated to inform the appropriate motor responses. Behaviorally relevant information must be kept in short-term memory. Thus the process of motor planning taps into multiple aspects of flexible behavior and offers a way of studying the mental processes that ultimately culminate in a movement, hence a window into cognition.

A neural correlate of motor planning was first reported in humans as a deflection in EEG recordings, over motor and parietal cortex, which anticipates voluntary movements (a.k.a. Bereitschaftspotential; Deecke et al. 1976). The EEG signal appears long before the onset of the movement and hundreds of milliseconds before

N. Li (✉) • Z.V. Guo • T.-W. Chen • K. Svoboda
Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA
e-mail: lin@janelia.hhmi.org

the subjects are aware of their desire to move (Libet 1985). The neural correlates of motor planning were discovered in the primate motor cortex by Tanji and Evarts (1976), who described neurons that discharged persistently before an instructed movement. This persistent activity ramped up shortly after the instruction, long before the movement onset, and predicted specific types of future movements. These findings opened the possibility of studying the mechanisms of motor planning at the level of neural circuits (Riehle and Requin 1989; Crutcher and Alexander 1990; Turner and DeLong 2000; Shenoy et al. 2013).

Behavioral paradigms in rodents are rapidly developing, and it is possible to train mice in behavioral tasks that dissociate planning and movement in time, analogous to the tasks used in primates (Guo et al. 2014a, b). The mouse is a genetically tractable organism, providing access to defined cell types for recordings and perturbations (Luo et al. 2008; O'Connor et al. 2009). In addition, the lissencephalic macrostructure of the mouse brain allows unobstructed access to a large fraction of the brain for functional analysis. We study motor planning in the context of a tactile decision behavior (Guo et al. 2014a; Li et al. 2015). Mice measure the location of an object using their whiskers and report their judgment by directional licking. We delineate when and how activity in specific cortical regions drives the tactile decision behavior in mice. New recording and perturbation methods are beginning to reveal the circuit mechanisms underlying motor planning that, in turn, will shed light on the biophysics of flexible behavior.

Head-Fixed Tactile Decision Behavior and Involved Cortical Regions

We developed a tactile decision task to track the flow of information in cortex during motor planning (Guo et al. 2014a; Li et al. 2015). Head-fixed mice measured the location of a pole using their whiskers and reported their decisions about object location with directional licking (Fig. 1a). In each trial, the pole was presented in one of two positions (anterior or posterior) during a sample epoch (1.3 s; Fig. 1b). During a subsequent delay epoch (1.3 s), the mice planned the upcoming response. An auditory go cue (0.1 s) signaled the beginning of the response epoch, when the mice reported the perceived pole position by licking one of two lickports (posterior → “lick right”, anterior → “lick left”). The mice achieved high levels of performance (~80 % correct). Responses before the go cue were rare (~13 %). They performed the tactile decision behavior across many sessions (up to 85 sessions, >400 trials per session), yielding tens of thousands of trials per mouse. The large number of trials allowed us to use optogenetic silencing to identify the cortical regions involved in an unbiased manner on a cortex-wide scale.

To achieve powerful cortical inactivation, we used transgenic mice expressing Channelrhodopsin-2 (ChR2) in cortical GABAergic interneurons. This photoinhibition silenced millimeter-size tissue volumes by photostimulating

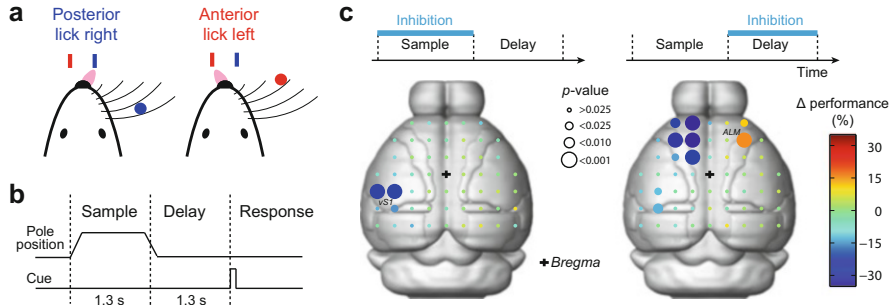


Fig. 1 Mapping the cortical regions underlying tactile decision behavior. **(a)** Head-fixed mouse responding “lick right” or “lick left” based on pole location. **(b)** The pole was within reach during the sample epoch. Mice responded with licking after a delay and an auditory go cue. **(c)** Fifty-five cortical locations were tested in loss-of-function experiments during different behavioral epochs. *Top*, photoinhibition during sample (*left*) and delay (*right*) epochs. *Bottom*, cortical regions involved in the tactile decision behavior during sample (*left*) and delay (*right*) epochs in “lick right” trials. *Color codes* for the change in performance (%) under photoinhibition relative to control performance. *Circle size codes* for significance (p values, from small to large; >0.025, <0.025, <0.01, <0.001). Figure adapted from Guo et al. (2014a)

through the intact skull (85 % activity reduction) with time resolutions on the order of 100 ms. We developed a scanning laser system to survey the neocortex for regions driving behavior during specific behavioral epochs. First, we outfitted the mice with a clear-skull cap preparation that provided optical access to half of the neocortex. A scanning system targeted photostimuli in a random access manner. Head-fixation and precise control of the laser position allowed each mouse to be tested repeatedly across multiple behavioral sessions. We tested 55 evenly spaced cortical volumes in sensory, motor, and parietal cortex for their involvement in the behavior by applying photoinhibition during specific behavioral epochs (Fig. 1c).

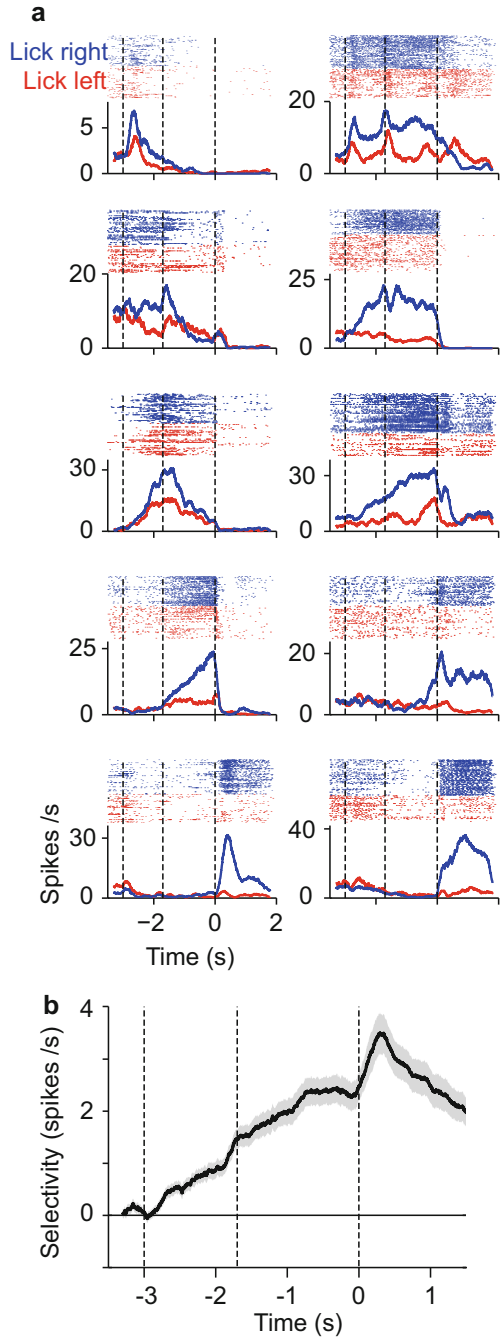
Inactivation of most cortical volumes did not cause any behavioral change. Inactivating vibrissal primary somatosensory cortex (vS1, “barrel cortex”) caused deficits in object location discrimination. The effect was temporally specific: inactivation during the delay epoch produced a much smaller deficit, suggesting that tactile information was transferred out of the vS1 during the sample epoch (Fig. 1c). During the delay epoch, preceding the motor response, inactivation of an anterior lateral region of the motor cortex (ALM) biased the upcoming movement (Fig. 1c). We used silicon probes to record single units from vS1 and ALM in mice performing the tactile decision behavior. Single unit recordings supported the photoinhibition experiments: a large fraction of neurons in vS1 showed object location-dependent activity during the sample epoch, whereas the majority of neurons in ALM showed movement-specific preparatory activity and peri-movement during the delay and response epochs. These results begin to outline the information flow in mouse cortex involved in the tactile decision behavior. The information flow is largely consistent with a serial scheme, where information is passed from sensory areas to motor areas during motor planning (Guo et al. 2014a).

Premotor Dynamics Underlying Motor Planning

A large fraction ($>70\%$) of ALM neurons significantly distinguished upcoming movements during at least one of the trial epochs ($p < 0.05$, two-tailed t-test; Fig. 2a). On error trials, when mice licked in the opposite direction to the instruction provided by object location (Fig. 1a), most ALM neuron activities predicted the licking direction (Guo et al. 2014a; Li et al. 2015). Such movement-specific activity is consistent with a role in planning and driving movements (Fig. 1c). Interestingly, the preparatory activity in ALM was not static but evolved with complex dynamics: subsets of neurons showed selective activity during the sample epoch whereas other neurons showed “bumps” of activity during different times of the delay epoch (Fig. 2a). Despite these fluctuating responses at the level of the single neurons, selectivity for the upcoming movement remained stable at the level of the population (Fig. 2b). On average, selectivity emerged in the sample epoch and ramped up throughout the delay epoch, reaching a maximum at the beginning of the response epoch (Fig. 2b). This ramping activity in ALM during the delay epoch is similar to the ramping activity reported in frontal (Hanes and Schall 1996; Murakami et al. 2014), parietal (Roitman and Shadlen 2002; Maimon and Assad 2006; Harvey et al. 2012), and motor cortex (Erlich et al. 2011; Thura and Cisek 2014), anticipating voluntary movements in primates and rodents. The ALM preparatory activity could provide a substrate for the maintenance of the motor plan during the delay epoch. Information about the upcoming movement is likely coded at the level of the population (Laurent 2002; Harvey et al. 2012; Shenoy et al. 2013; Murakami and Mainen 2015). Consistent with a distributed code, we found that spatially intermingled ALM neurons in each hemisphere have a preference for either contra- or ipsi-lateral movements in roughly equal proportions (Li et al. 2015).

The preparatory activity encodes upcoming movement, yet it does not cause movement during planning. How does preparatory activity evolve into commands that descend to motor centers to trigger movement? Paradoxically, ALM neurons in each hemisphere have a preference for contra- or ipsi-lateral movements in roughly equal proportions, yet unilateral inactivation of ALM biased the upcoming movement to the ipsi-lateral direction. How does silencing a brain area with non-lateralized selectivity cause a directional movement bias? We measured neuronal activity within hierarchically organized ALM circuits. ALM projection neurons include two major classes: intratelencephalic (IT) neurons that project to other cortical areas and pyramidal tract (PT) neurons that project out of the cortex, including to motor-related areas in the brainstem (Komiyama et al. 2010; Shepherd 2013). IT neurons connect to other IT neurons and excite PT neurons, but not vice versa. PT neurons are thus at the output end of the local ALM circuit (Morishima and Kawaguchi 2006; Brown and Hestrin 2009; Kiritani et al. 2012; Shepherd 2013). We recorded activity from identified IT and PT neurons using cell-type specific electrophysiology and two photon imaging. We found that ALM IT neurons have mixed preparatory activity for both ipsi- and contra-lateral movements. Contra-lateral population activity in PT neurons arose late during the delay

Fig. 2 Head-fixed tactile decision behavior and involved cortical regions. **(a)** Nine examples of ALM neurons. *Top*, spike raster. *Bottom*, PSTH. Correct “lick right” (*blue*) and “lick left” (*red*) trials only. *Dashed lines*, behavioral epochs. **(b)** Average ALM population selectivity in spike rate (\pm s.e.m. across neurons, bootstrap). Selectivity is the difference in spike rate between the preferred and non-preferred movements. Time zero is the onset of the go cue. Averaging window, 200 ms



epoch to drive directional licking. To test the causal role of the PT neuron population activity in driving movements, we manipulated PT neurons by expressing ChR2 in mouse lines selectively expressing cre in these neurons. Weak activation of the PT neurons during movement planning could “write-in” specific motor plans that resulted in contra-lateral licking movements. These results suggest that, during movement planning, distributed preparatory activity in IT neuron networks is converted into a movement command in PT neurons (‘output-potent’ activity; Kaufman et al. 2014), which ultimately triggers directional movements (Li et al. 2015).

Open Questions

Several key questions remain unsolved. What circuit mechanisms are responsible for the maintenance of motor plan during short-term memory? How is sensory information integrated into the motor plan? How do the basal ganglia and motor thalamus interact with cortical regions during motor planning? Answering these questions will require recordings and manipulation of specific cell types. Importantly, architectural and cell type information must be incorporated into models of cortical dynamics. Tools to manipulate projections between brain regions are needed to study the interactions between brain regions. Finally, there is still a long way to go in developing richer behavioral paradigms that tap into the capabilities of the mammalian brain.

Acknowledgment This work was supported by Howard Hughes Medical Institute. N.L. is a Helen Hay Whitney Foundation postdoctoral fellow.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work’s Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work’s Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Brown SP, Hestrin S (2009) Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature* 457:1133–1136
- Crutcher MD, Alexander GE (1990) Movement-related neuronal activity selectively coding either direction or muscle pattern in three motor areas of the monkey. *J Neurophysiol* 64:151–163

- Deecke L, Grozinger B, Kornhuber HH (1976) Voluntary finger movement in man: cerebral potentials and theory. *Biol Cybern* 23:99–119
- Erlich JC, Bialek M, Brody CD (2011) A cortical substrate for memory-guided orienting in the rat. *Neuron* 72:330–343
- Guo ZV, Li N, Huber D, Ophir E, Gutnisky DA, Ting JT, Feng G, Svoboda K (2014a) Flow of cortical activity underlying a tactile decision in mice. *Neuron* 81:179–194
- Guo ZV, Hires SA, Li N, O'Connor DH, Komiyama T, Ophir E, Huber D, Bonardi C, Morandell K, Gutnisky D, Peron S, Xu NL, Cox J, Svoboda K (2014b) Procedures for behavioral experiments in head-fixed mice. *PLoS One* 9:e88678
- Hanes DP, Schall JD (1996) Neural control of voluntary movement initiation. *Science* 274:427–430
- Harvey CD, Coen P, Tank DW (2012) Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* 484:62–68
- Kaufman MT, Churchland MM, Ryu SI, Shenoy KV (2014) Cortical activity in the null space: permitting preparation without movement. *Nat Neurosci* 17:440–448
- Kiritani T, Wickersham IR, Seung HS, Shepherd GM (2012) Hierarchical connectivity and connection-specific dynamics in the corticospinal-corticostriatal microcircuit in mouse motor cortex. *J Neurosci* 32:4992–5001
- Komiyama T, Sato TR, O'Connor DH, Zhang YX, Huber D, Hooks BM, Gabitto M, Svoboda K (2010) Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature* 464:1182–1186
- Laurent G (2002) Olfactory network dynamics and the coding of multidimensional signals. *Nat Rev Neurosci* 3:884–895
- Li N, Chen TW, Guo ZV, Gerfen CR, Svoboda K (2015) A motor cortex circuit for motor planning and movement. *Nature* 519:51–56
- Libet B (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav Brain Sci* 8:529–539
- Luo L, Callaway EM, Svoboda K (2008) Genetic dissection of neural circuits. *Neuron* 57:634–660
- Maimon G, Assad JA (2006) A cognitive signal for the proactive timing of action in macaque LIP. *Nat Neurosci* 9:948–955
- Morishima M, Kawaguchi Y (2006) Recurrent connection patterns of corticostriatal pyramidal cells in frontal cortex. *J Neurosci* 26:4394–4405
- Murakami M, Mainen ZF (2015) Preparing and selecting actions with neural populations: toward cortical circuit mechanisms. *Curr Opin Neurobiol* 33C:40–46
- Murakami M, Vicente MI, Costa GM, Mainen ZF (2014) Neural antecedents of self-initiated actions in secondary motor cortex. *Nat Neurosci* 17:1574–1582
- O'Connor DH, Huber D, Svoboda K (2009) Reverse engineering the mouse brain. *Nature* 461:923–929
- Riehle A, Requin J (1989) Monkey primary motor and premotor cortex: single-cell activity related to prior information about direction and extent of an intended movement. *J Neurophysiol* 61:534–549
- Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22:9475–9489
- Shenoy KV, Sahani M, Churchland MM (2013) Cortical control of arm movements: a dynamical systems perspective. *Annu Rev Neurosci* 36:337–359
- Shepherd GM (2013) Corticostriatal connectivity and its role in disease. *Nat Rev Neurosci* 14:278–291
- Tanji J, Evarts EV (1976) Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. *J Neurophysiol* 39:1062–1068
- Thura D, Cisek P (2014) Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making. *Neuron* 81:1401–1416
- Turner RS, DeLong MR (2000) Corticostriatal activity in primary motor cortex of the macaque. *J Neurosci* 20:7096–7108

The Visual Brain: Computing Through Multiscale Complexity

Yves Frégnac, Julien Fournier, Florian Gérard-Mercier, Cyril Monier, Marc Pananceau, Pedro Carelli, and Xoana Troncoso

Abstract Information coding in sensory neurons is both digital, in terms of neuronal output spike timing and rate, and analog, produced by the irregular subthreshold changes in somatic and dendritic membrane potential resulting from synchronized volleys of synaptic inputs. Intracellular recordings give a unique access to a composite multiscale signal where the local microscopic integration process realized by a single neuron can be studied in the global mesoscopic context of the “unseen” units afferent to the recorded cell. This chapter shows how reverse engineering approaches can be used in the primary visual cortex of higher mammals to reveal the hidden complexity of visual processing and establish causal links between the functional dynamics of synaptic echoes in primary visual cortex and perceptual biases in low-level, non-attentive perception.

Introduction

A still highly controversial issue in visual perception concerns the definition of receptive fields (RFs). Do they represent a static transfer function associated with each sensory neuron or a dynamic integration window open on the sensory world and continuously reshaped by contextual changes in the input statistics? How do RF properties adapt to contrast/luminance non-stationarities experienced during the oculomotor exploration of the visual world? Our understanding of early visual processing (implemented in our brain by the retino-thalamo-cortical projection)

Y. Frégnac (✉) • F. Gérard-Mercier • C. Monier • M. Pananceau • P. Carelli • X. Troncoso
Centre National de la Recherche Scientifique (CNRS-UNIC), Unité de Neuroscience,
Information et Complexité (UNIC), Gif-sur-Yvette, France
e-mail: fregnac@unic.cnrs-gif.fr

J. Fournier
Centre National de la Recherche Scientifique (CNRS-UNIC), Unité de Neuroscience,
Information et Complexité (UNIC), Gif-sur-Yvette, France

Max Planck Institute for Brain Research, Max von Laue Str. 4, 60438 Frankfurt am Main,
Germany

in higher mammals has been shaped, to a large degree, with a hypothesis-driven conception of sensory architecture in mind. This bias is best illustrated by the 50-year dominance of the hierarchical schema of genesis of RF specificity (Hubel and Wiesel 1962, 1968, 2005), where visual attribute extraction in the early visual system results from a serial processing starting from non-oriented retinal and thalamic RFs to a cascade of oriented edge detectors of increasing complexity across multiple successive relays (thalamus, primary and secondary cortical areas). This feedforward view of visual processing is based on the repetition, at each stage of integration, of canonical, but highly specific, rules of anatomical convergence from which the function derives. Consequently, our current understanding of RF genesis is largely bottom-up driven and contradicts Aristotle's principle that "the whole is greater than the sum of its parts." Although this simplifying view of sensory processing has led to major advances [review in Alonso (2002); see also the remarkable obituary tribute to David Hubel by Kevan Martin (2014)], it fails to account for the functional complexity expected from the recurrent structural connectivity of cortical subcircuits (Douglas and Martin 2004) on the one hand and the non-linear nature of the dynamic interactions between excitation and inhibition during sensory processing (Borg-Graham et al. 1998; Monier et al. 2008). Furthermore, as emphasized by Bruno Olshausen (2014), RF classification has been established using highly standardized and parametrized sensory contexts (spots, bars and gratings), which have little to do with the rich spatio-temporal statistics experienced during the natural scene-viewing conditions of our everyday life. The principle of maximization of a single neuron's firing rate, which initially guided the "neuronal doctrine" (Barlow 1972), does not hold any more when RFs are engaged by non-optimal stimuli most of the time, leading, for natural scene processing, to a sparser and unexpectedly temporally precise spiking regime in the primary visual cortical area (Vinje and Gallant 2000; Baudot et al. 2013). A third conceptual limit to our present knowledge of early visual processing is that most modeling efforts have been targeted at explaining sensory discharges only at the spike level in a purely phenomenological perspective (see Carandini et al. 2005 for a review) rather than aiming at elucidating causal, conductance-based mechanisms regulating the temporal selectivity of the spiking opportunity window (Haider et al. 2010; Baudot et al. 2013).

Although *in vivo* intracellular recordings are difficult to practice in higher mammals, and thus often limited to the anesthetized and paralyzed preparation (but see Tan et al. 2014 for a "tour de force" in the behaving non human primate), they still appear to be the technique of choice to address quantitatively the synaptic nature of the RF (Fig. 1a, b). In contrast to most imaging methods (voltage-sensitive dye, two-photon) or correlation studies based on multiple recordings, intracellular recordings allow us to differentiate the local microscopic integration process achieved by the single neuron from the more mesoscopic contribution of the network of unseen units that influence at each point in time the activity of the recorded cell (Fig. 1c).

We review here recent intracellular electrophysiological studies from the Frégnac lab (CNRS-UNIC), done in the visual cortex of the anesthetized

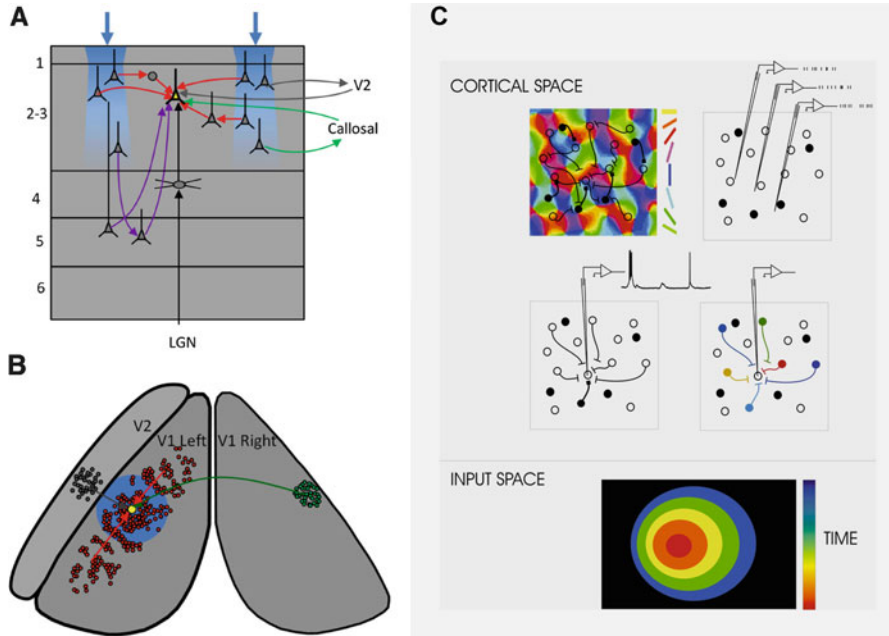


Fig. 1 Bridging microscopic and mesoscopic scales. **(a)** The synaptic RF (adapted from Huang et al. 2014). Sensory cortical neurons integrate the feedforward drive from the thalamus (LGN), eventually relayed by intracolumnar connections and amplified by recurrent local connectivity, with the lateral input provided by intrinsic, horizontal long-distance connections and cortico-cortical feedback and interhemispheric callosal loops. **(b)** Retrieving mesoscopic dynamics from intracellular recordings (adapted from Frégnac et al. 2007). *Top row*, classical methods for studying evoked sensory dynamics: *left*, voltage-sensitive, dye-imaging map based on hemodynamic signals; *right*, same network analyzed with simultaneous multiple electrode recordings (blind connectivity). *Middle row*, intracellular recording where reverse engineering methods allow extraction of the “effective connectivity,” influencing the membrane potential at any point in time. Synaptic functional imaging, based on feature selectivity and space and time, allows us to identify the synaptic sources and reconstruct predictions of the full network dynamics (*bottom row*)

mammalian brain and addressing separately two scales of spatial integration in the V1 RF: (1) the inner ON-OFF organization of the RF core (underlying the “Simple” vs. “Complex” typology), depending on the balance between feedforward and local recurrent connectivity, and (2) the “association field” extending in the “silent” surround of the RF, from which subthreshold activation can be evoked through the propagation along long-distance, slowly conducting “horizontal” connections intrinsic to V1 (Binguier et al. 1999; Chavane et al. 2011; Frégnac 2012; Gérard-Mercier et al. in preparation).

The claim we make here is that the intracellular subthreshold membrane potential signal gives unique access to the multiscale nature of cortical processing and that reverse engineering methods can be designed to unfold, from the intracellular reading of synaptic echoes, the mesoscopic dynamics of the afferent network

(Fig. 1, right panel). In the first part of this review, we will show that the hidden complexity revealed by this approach demonstrates how limited our current understanding is of the bottom-up **emergence** of dynamic properties in visual RFs in the early visual system. Furthermore, it unravels the existence of **immergence** processes through which the collective mesoscopic constraints imposed by the distributed sensory input regulate the functional expression of individual RF properties in a top-down fashion. In the second part, we will illustrate how the decoding of synaptic echoes originating from the silent surround of the RF allows us, in a surprising way, to extract functional structural biases that may serve the self-organization of psychological Gestalt laws in the non-attentive brain. These last findings can be seen as one of the few successful attempts to link visually evoked synaptic dynamics to perceptual biases and low-level perception, thus establishing a **causal bridge between microscopic and macroscopic scales**.

The Synaptic Imprint of Mesoscopic “Immergence” in Visual RFs

RFs in primary visual cortex (V1) are categorized as “Simple” or “Complex,” depending on their spatial selectivity to stimulus contrast polarity, leading to the retinotopic segregation (Simple) or overlap (Complex) of “ON” and “OFF” sub-fields (Hubel and Wiesel 1962, 1968). Since the pioneering work of Hubel and Wiesel in the cat and macaque visual cortex, most extracellular surveys have reported V1 RFs with intermediate behaviors (Henry 1977). Intracellular recordings have shown that the separation between Simple and Complex RFs hides a continuous distribution of synaptic inputs, in particular with regard to their degree of linearity (in the sense of system theory; Skottun et al. 1991; Priebe et al. 2004). In spite of the general acceptance that Simple and Complex RF organizations correspond to distinct balance levels between linear and nonlinear contributions at the synaptic level, few studies have investigated whether the functional expression of the Simple or Complex nature of V1 RFs depends, in the same cell, on the spatiotemporal statistics of the stimulus. This may be surprising, since gain-control mechanisms are known to ensure the contrast invariance of the cell selectivity, and it is well established that adaptation to stimulus contrast does not have the same effect on the Simple-like and Complex-like components of V1 cell responses when assessed with drifting gratings. Still, most previous studies of RF adaptation to visual statistics focused either on linear RF components in Simple cells or on nonlinear components in Complex cells and not on the differential adaptation of these two components in the same RF (but see Yeh et al. 2009).

In a series of studies targeting the subthreshold definition of the RF, we realized that RF maps varied significantly between stimulus conditions, leading us to reassess in a quantitative way how the Simple/Complex nature of individual V1 cells depends on the full field visual context used to probe the RF (Fournier

et al. 2011, 2014). This was done by comparing systematically, in the same cell recorded intracellularly, the synaptic responses to three classical RF mapping protocols based on white noise: sparse noise, ternary dense noise and flashed Gabor noise. A surprising result was that the linear kernel estimate differed between these various contextual noises, in contrast with the prediction of invariance made by cascade L-N-P models of V1 RFs (according to the so-called Bussgang theorem; Bai et al. 2007). Intracellular recordings revealed that, for most V1 cells, there was no such thing as an invariant RF type, but that the relative weights of Simple-like and Complex-like RF components were scaled such as to make the same RF more Simple-like with dense noise stimulation and more Complex-like with sparse or Gabor noise stimulations (example in Fig. 2a; population analysis in Fig. 2b). However, once these context-dependent RFs were convolved with the corresponding stimulus, the balance between Simple-like and Complex-like contributions—in terms of input current—to the synaptic responses appeared to be invariant across input statistics (Fig. 2c; Fournier et al. 2011).

This invariance of the ratio between the linear/nonlinear input current contributions suggests a novel form of homeostatic control of V1 functional properties, where the expressed network nonlinearities are optimized by the statistical structure of the visual input. This study is the first, to the best of our knowledge, to show such clear changes in terms of spatiotemporal reorganizations of synaptic and discharge fields at the single cell level, interpretable as a coherent adaptive behavior at the cortical population level. The claim made here is that these effects are more detectable at the subthreshold than at the spiking level, where additional static non-linearities may interfere with the global read-out of the connectivity adaptation rule.

A functional interpretation of these data could be that the Simple or Complex nature of V1 RFs arises from a variable balance between feed-forward and lateral inputs, with the feed-forward drive providing the Simple-like component whereas the recurrent lateral connections would convey Complex-like contributions (Fig. 3, left). Accordingly, the results might be explained by the functional recruitment of lateral interactions in sparse stimulation conditions and by the decoupling of adjacent cortical columns in dense visual contexts. This view is supported by other studies, realized for instance by the group of Matteo Carandini, suggesting that the lateral propagation of activity between adjacent cortical units decreases substantially when the stimulus contrast is increased (Fig. 3, right, adapted from Nauhaus et al. 2009). In view of these different results, the stimulus dependence of the lateral cortical interactions likely generalizes to other stimulus dimensions, rather than remaining exclusive to the local contrast. Similar effects might be obtained by increasing the spatial or temporal density of the stimulus, with the important parameter probably being the effective contrast along the stimulus feature dimensions for which the cell is selective.

To enrich the predictive power of the synaptic RF model, we decomposed the second-order kernel estimate obtained by a truncated Volterra expansion of the membrane potential response to dense noise, into a non-linear combination of parallel Simple-like filters in a way similar to the spike-triggered covariance (STC) introduced by the groups of Simoncelli and Movshon (Rust et al. 2005).

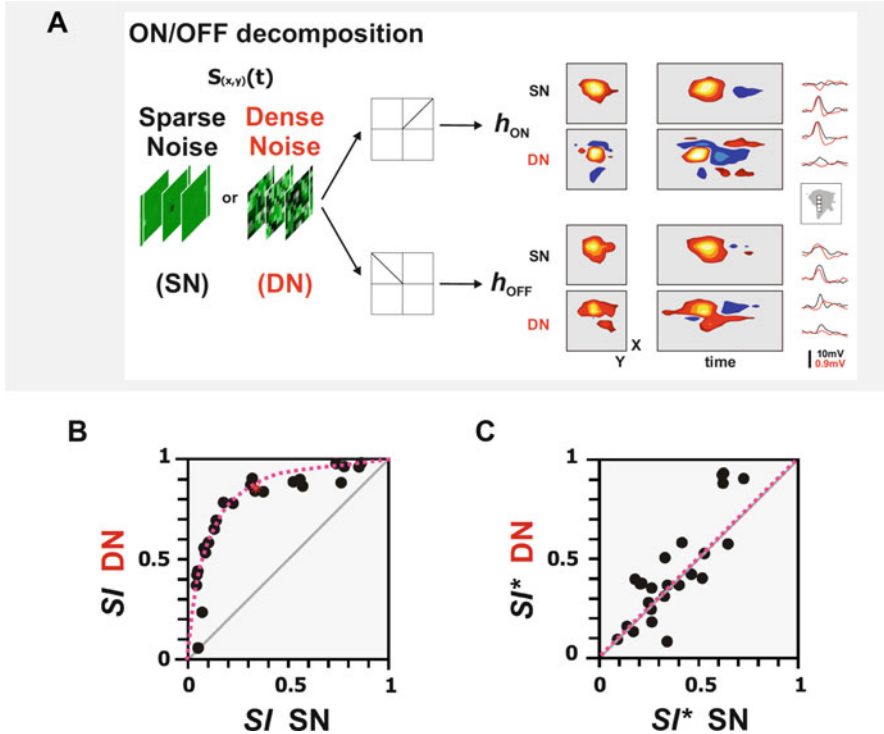


Fig. 2 The functional expression of V1 RFs depends on input statistics. **(a)** Example of a layer 2–3 cell: the ON and OFF kernels are shown for the two noise input statistics used to map the subthreshold RF (*SN* sparse noise, *DN* dense noise). The *shaded boxes* represent the X-Y and X-time features of the RF filter, with ON and OFF subfields represented in *red* and *blue*, respectively. Note that the maps are Simple-like for the SN statistics [spatially segregated ON and OFF subfields (X-Y map) and reversal of the spatio-temporal filter polarity with time (X-t profile)] and Complex-like for the DN statistics (spatially overlapping subfields). *Right column*, the individual kernel waveforms (mV), detailed for four different pixels (*inset*), are represented in *red* for DN and *black* for SN. Note the divisive effect of dense noise compared to sparse noise on the kernel estimate amplitude (by a tenfold factor). **(b, c)** Population analysis of the Stimpleness Index (given by the ratio of the linear kernel energy divided by the total RF kernel energy). “0” stands for Complex RFs (purely non-linear) and “1” for Simple RFs (purely linear). **(b)** Population bi-histogram plot linking (on a cell-by-cell basis) the Stimpleness Index (SI) for sparse noise (SN, abscissa) and dense noise (DN, ordinate) stimulation. Hyperbolic fit as a *pink dotted curve*. **(c)** Same population bi-histogram for SI* values obtained after convolution of the kernels with the visual input waveform. Note the realignment of the points (cells) along the identity line (bihistogram diagonal). See details in Fournier et al. (2011)

Although the STC method was applied with success at the spiking level to reveal non-linear subunits in V1 and MT cells in the macaque, it failed to reveal more diversity in the cat (Touryan et al. 2002, 2005), probably for technical reasons linked to the limited number of spikes. This potential problem is bypassed here by applying similar techniques to the continuous intracellular membrane potential

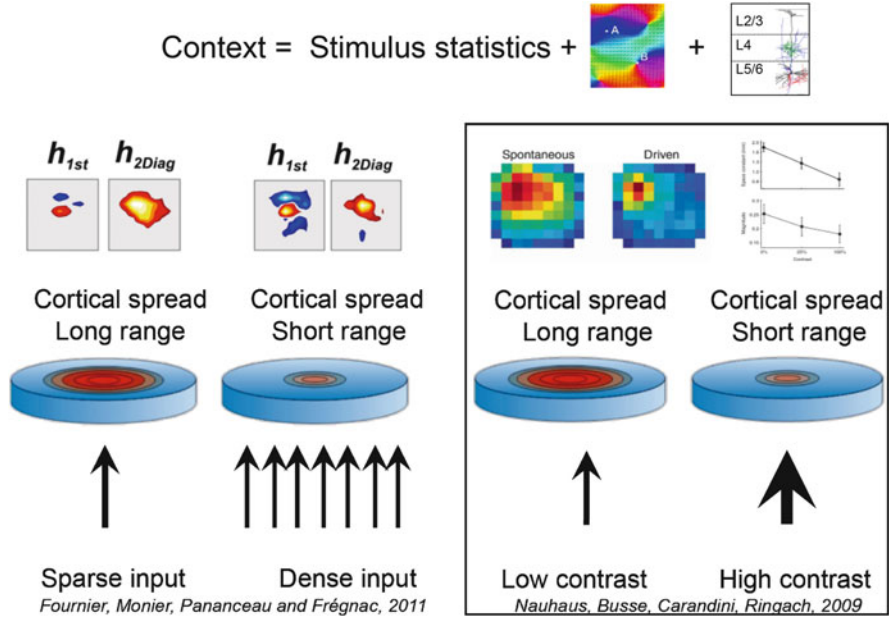


Fig. 3 Emergence of mesoscopic network organization on microscopic response properties. Changes in the mesoscopic context are provided by, functionally, non-stationarities in the sensory input drive statistics and, structurally, by non-homogeneities in the columnar and lateral network neighborhood (Monier et al. 2003). (A) Homeostasis Rule as a function of spatial and temporal density (Fournier et al. 2011): regulation of the functional expression of the RF (h_{1st} for linear and h_{2nd} Diagonal for the diagonal term of the second-order kernel) by input statistics (*left*, SN; *right*, DN). The spread of lateral activation is wider for SN than for DN. (B) Homeostasis Rule as a function of contrast level (adapted from Nauhaus et al. 2009): similar regulation of the lateral spread through horizontal connectivity by the stimulus contrast

signal, resulting in a much higher sensitivity for detecting stimulus-evoked sub-threshold events contributing to the spike discharge (Fournier et al. 2014). Conductance or current-based principal component analysis (PCA) allows us to decompose the Simple-like and Complex-like components of the subthreshold RF into a parallel set of functionally distinct subunits, consisting of a Simple-like sub-unit whose linear output accounts for the push-pull components of the RF and a variable number of Complex-like sub-units that contribute in a fully rectified manner to the cell response (Fig. 4, left). Results show that both Simple and Complex RFs exhibit a remarkable diversity of excitatory and inhibitory Complex-like contributions, which differ in orientation and spatial frequency selectivity from the linear RF, even in layer 4 and layer 6 Simple cells that received direct input from the thalamus (exemplified in Fig. 4, right). Furthermore, the diversity of Complex-like contributions recovered at the subthreshold level is expressed in the cell spiking output.

These results demonstrate that the Simple or Complex nature of V1 RFs does not rely on the diversity of Complex-like components received by the cell from its

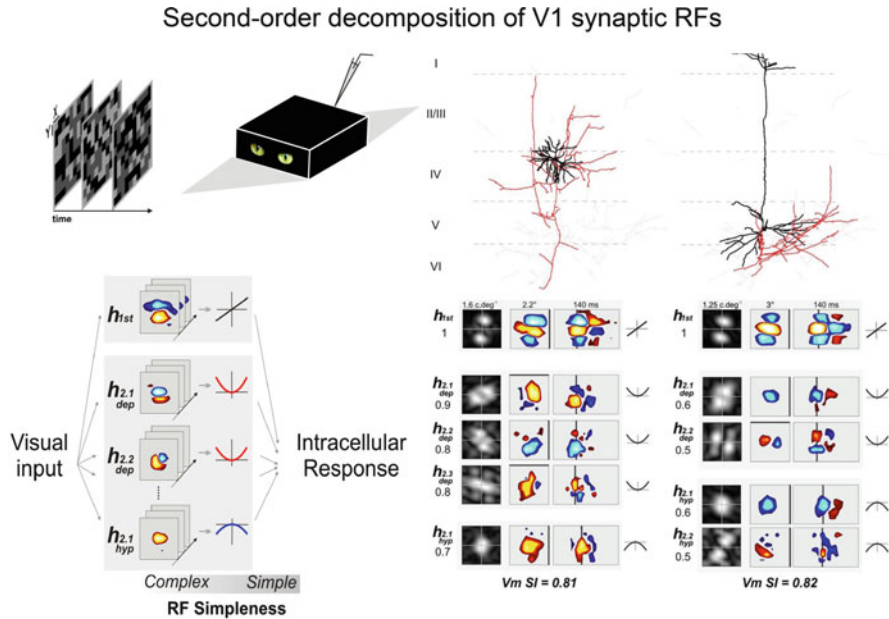


Fig. 4 Filter bank decomposition of the subthreshold V1 RF (adapted from Fournier et al. 2014). *Left*: decomposition principle: each branch of the filter bank is composed of a Simple-like filter followed by an identity contrast function (linear kernel, *upper branch*) and by a parallel bank of linear subunits feeding excitatory (*red*) and inhibitory (*blue*) quadratic contrast-dependent non-linearities (*lower parallel branches*). *Right*: example of RF decomposition for two biocytin reconstructed cells in, respectively, layer 4 (*middle*) and at the border between layers 5/6 (*right*) in cat visual subunit. Each subunit weight (in the decomposition) is given below each kernel component

synaptic afferents but on the relative imbalance between the weights of the Simple-like and Complex-like synaptic contributions. In spite of the likelihood that the Simple-like RF subunit results from the push-pull arrangement of excitatory and inhibitory feedforward inputs selective for the same orientation, the diversity of feature selectivity expressed by the Complex-like RF subunits is not consistent with a strict iso-orientation preference rule for excitatory and inhibitory input conductance as generally posited (Ferster and Miller 2000; Priebe and Ferster 2012). Although the estimated Complex-like subunits are operational filters that do not necessarily correspond to the RFs of neurons presynaptic to the recorded cell, they bear a striking resemblance to the linear RF of V1 Simple cells, which suggests that they could correspond to separate subcircuits originating from within the cortex (Rust et al. 2005; Chen et al. 2007). The diversity of orientation and spatial frequency preferences of the Complex subunits agrees with that found in the tuning of the excitatory and inhibitory input conductances measured by voltage clamp techniques *in vivo* and previously reported by our lab (Monier et al. 2003, 2008). Taken together, these intracellular results support the hypothesis that the Complex-like components of V1 RFs arise from lateral interactions between adjacent cortical

columns and are consistent with the proposal that the Simple or Complex nature of V1 RFs arises from the respective balance between feedforward and lateral connectivity (Chance et al. 1999; Tao et al. 2004). This wide functional spectrum of Complex-like synaptic contributions to both Simple and Complex RFs may constitute the skeleton of a multi-competent substrate allowing V1 cells to adapt on-the-fly to the abrupt changes in the spatio-temporal statistics of visual inputs (Fig. 4, right).

Synaptic Correlates of Perceptual Gestalt Laws in the “Silent Surround” of V1 RFs

The synaptic RF stems from the interplay of distinct sets of connections, the feedforward drive from the thalamus relayed eventually by vertical processes within the cortical column, the local recurrent reverberation usually confined within a hypercolumn, the long-distance connectivity intrinsic to V1 (that may even originate from the other hemisphere through the corpus callosum) and the feedback from higher cortical areas (Fig. 1a, b). The cat and ferret visual cortex appear to be ideal experimental models to study horizontal connectivity (Kisvarday et al. 1997; Bosking et al. 1997), since many reconstructed axons of pyramidal cells remaining within the gray matter have been shown to extend over several hypercolumns (up to 6–8 mm in the cat; Kisvarday et al. 1997; Callaway and Katz 1990; Gilbert and Wiesel 1983; Gilbert and Li 2012; Buzàs et al. 2006; but see Martin 2014). In spite of some pioneering attempts (Kasamatsu et al. 2010; Mizobe et al. 2001), only limited physiological data have addressed the synaptic contribution of the “silent” surround of the classical V1 RFs, from which impulse-like stimuli fail to evoke a spiking response. Consequently, the role of long distance horizontal connectivity in influencing the response gain within the classical RF, and in particular in boosting it for specific center-surround stimulus conditions (Jones et al. 1980; Sillito et al. 1995; Sillito and Jones 1996), remains an issue of debate. In spite of this uncertain status, horizontal connectivity has long been presented as the biological substrate of iso-preference binding in the electrophysiological and psychophysical cortical literature (review in Gilbert and Li 2012; Frégnac and Bathellier 2015). This principle was derived from a developmental rule that posited that “who fires together (or is alike) tend to wire together” (Callaway and Katz 1990). At the psychophysical level, this view corresponds to the perceptual “association field” concept, developed by Field, Hess and their colleagues in the 1990s (Field et al. 1993). This concept assumes the instantaneous induction of collinear and, to a lesser extent, co-circular facilitation by the static presentation of oriented contrast edges. This elegant psychophysical hypothesis accounts in humans for the “pop-out” perception of smooth contiguous path integration even when immersed in a sea of randomly oriented edge elements (Fig. 5a, top; Field et al. 1993) and the facilitation of target detection by high contrast co-aligned flankers (Fig. 5a, bottom;

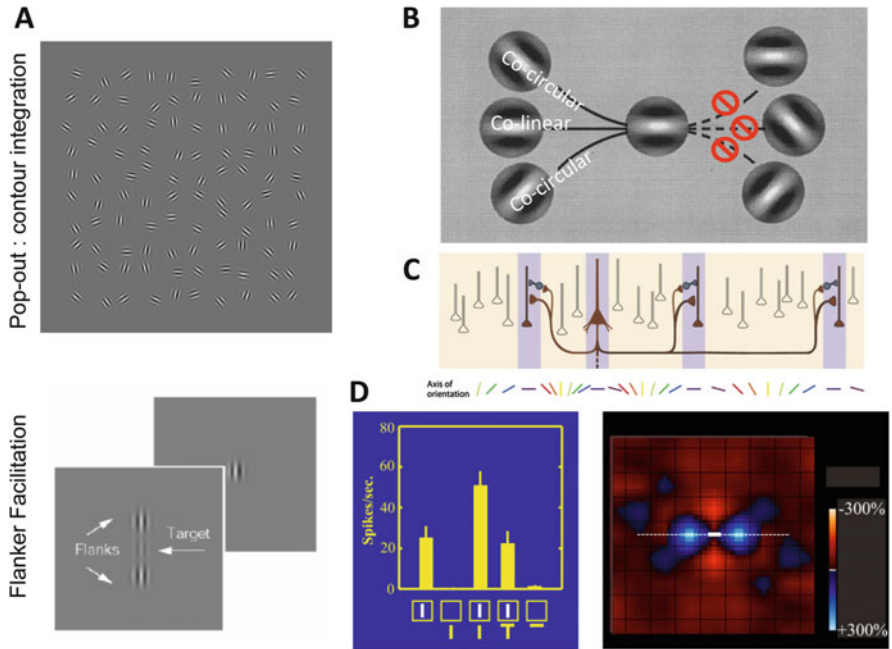


Fig. 5 The perceptual association field and its neuronal correlate in the attentive brain (reviewed in Frégnac and Bathellier 2015). (a) *Top*: “pop-out” emergence of a continuous integration path in a sea of randomly oriented Gabor patches (Field et al. 1993). *Bottom*: facilitation of detection of a low contrast vertical Gabor element induced by the simultaneous presentation of co-aligned high contrast flanker elements (Polat and Sagi 1993). (b) Hypothetical association field induced by an oriented element through lateral interactions promoting co-alignment and co-circularity (Field et al. 1993). (c) The “iso-functional binding” hypothesis (Gilbert and Li 2012). An individual superficial layer cortical pyramidal cell forms long-range connections that extend many millimeters parallel to the cortical surface. Long-range connections (>500 μm from the injection center) tend to link columns of similar orientation preference. (d) The “neural facilitation field” (Li et al. 2006). *Left*, the responses of V1 neurons are amplified in the awake behaving monkey by collinear contours extending outside the RF. Introducing a cross-oriented bar between the collinear segments blocks the contour-related facilitation. *Right*, two-dimensional map of facilitatory (blue) and inhibitory (red) modulation of the response to an optimally oriented line segment centered in the RF (horizontal white bar). The spiking modulation is suppressed by anesthesia

Polat and Sagi 1985). At the neuronal level, this view is supported by the peculiar anatomy of long-distance horizontal connections emitted by supragranular pyramidal cells found consistently in higher mammals (but see Martin et al. 2014) and the electrophysiological demonstration of a “neural facilitation field” (Fig. 5b, c; Gilbert and Li 2012). These latter experiments, realized in the attentive behaving monkey, demonstrated an impressive boosting of the response gain to an optimally oriented contrast edge within the classical RF when flankers were simultaneously flashed in the “silent surround” and co-aligned along the preferred orientation axis of the extracellularly recorded cell. Most remarkably, Charles Gilbert, Wu Li and his colleagues showed that, to be expressed, the co-linearity binding rule required

the existence of top-down signals, present in the target-attending monkey, since the effect was weakened by diverted attention (Li et al. 2006) and the ability to learn contour integration was suppressed by anesthesia (Li et al. 2008).

These previous studies provided, nevertheless, an indirect answer since they addressed only the modulatory nature of the center-surround effects, without probing the existence of a subthreshold influence. This issue has been addressed intracellularly in the anesthetized mammal, and our lab has demonstrated repeatedly, in the context of various stimulation protocols, the existence of long-distance propagation of visually evoked activity through lateral (and possibly feedback) connectivity outside the classical RF (Bringuier et al. 1999; Frégnac 2012; Gerard-Mercier et al. 2016; Troncoso et al. 2015). This propagation, initially hypothesized by Amiram Grinvald and inferred from the synaptic echoes we recorded intracellularly, has since been confirmed in the same species by voltage sensitive dye (VSD) imaging techniques (Benucci et al. 2007; Chavane et al. 2011), which provide a direct visualization of the horizontal propagation pattern at the mesoscopic level of the V1 retinotopic map. Most remarkably, the VSD waves were found to travel at the same speed as that inferred from intracellular recordings (0.3 m/s).

In a recent intracellular study (Gérard-Mercier et al. 2014, in preparation), we reinvestigated the association field concept to demonstrate whether a structure-functional bias might be still detected at the subthreshold level, even in the absence of attention-related signals. By averaging synaptic response properties in a unified “cellulo-centric” reference frame centered on the discharge field center and realigned with the spike-based orientation preference, we found a coherent spatial organization of visual synaptic responses, reflecting the grouping bias of the “perceptual association field” for collinear contours (Field et al. 1993). This result, apparently contradictory to Gilbert and Li’s failure to find the “facilitatory neural field” under anesthesia, is seen only at the population level by summation across cells. The most likely interpretation is that a mean-field effect (in the sense of physics) is needed to enhance a slight bias in the subthreshold impact of the synaptic connectivity intrinsic to V1. Its expression is revealed (or facilitated) here by the use of 3–4° test-oriented stimuli (Gabor patch) that recruit by spatial summation the whole extent of the aggregate RF of a hypercolumn in the cat. Our current working hypothesis is that a critical threshold of spatial synergy and temporal summation has to be trespassed to make the weak functional impact of these long-range interactions (in the mV range) detectable, as suggested from a prior combined VSD and intracellular study done in collaboration with the lab of Amiram Grinvald (Chavane et al. 2011). Preliminary intracellular data show that two- to six-stroke apparent motion (AM) sequences, riding in phase with horizontal activation in a centripetal way towards the RF center, are effective enough to unmask suprathreshold filling-in responses in the unstimulated RF core (Troncoso et al. 2015).

Our work provides, for the first time, intracellular evidence in the anesthetized mammal for synaptic correlates of low-level perception, closely dependent on the spatiotemporal features of the synaptic integration field of V1 neurons and most

likely linked to intra-V1 horizontal connectivity. These findings also agree with the concept of a “dynamic association field,” whose spatial anisotropy and extent are transiently updated and reconfigured as a function of changes in the retinal flow statistics imposed during visuomotor exploration of natural scenes (Frégnac 2012). According to this still hypothetical view, the propagation of intracortical depolarizing waves at the mesoscopic V1 map level would help in broadcasting an elementary form of **collective predictive “belief”** to distant parts of the network, at a time when they are not yet engaged by the stimulus drive. We propose that the in-phase association of horizontal and feedforward input could provide the synaptic substrate for implementing the psychological Gestalt principles of common fate and axial collinearity (review in Wagemans et al. 2012). On a more conjectural note, since a visual flow in the order of 100–250°/s in retinal space is needed to maintain—in cat V1—the feedforward flow in phase or slightly ahead of intra-V1 propagation, one may expect the amplification of visual responses for edges collinear to the motion path during specific phases of brisk eye-movements, namely saccadic exploration or large changes of gaze between distant fixation locations. This unexpected process could account for the observation of transient peaks of responses for fast-moving contours coaligned with the RF axis (Barry Richmond, personal communication; Judge et al. 1980) and the induction of filling-in responses for fast centripetal radial flow (Troncoso et al. 2015).

Conclusion

We conclude from this review that the functional complexity in the early visual system is largely underestimated and that the functional organization and preference expressed in visual cortical RFs result from the coordination by input statistics dynamics of overlaid activity processes operating at different spatial integration scales. We have illustrated here what insight can be possibly gained by the comparison between different levels of integration. Reverse engineering on intracellular and spiking signals shows that part of the “effective” connectivity contributing to the RF is missed/ignored when models and data collection are confined at the spiking level. Mapping of the hidden non-linearities in the subthreshold RF reveals unexpected **immurgence** processes, driven by the stimulus, through which the global activity control extending within and beyond the cortical hypercolumn regulates the functional expression of more microscopic properties, such as the apparent “Simpleness” of individual RFs. This feature can be seen as a **top-down** influence of the more mesoscopic levels of organization, typical of complex dynamic systems based on nested processing. The unfortunate consequence of this physiological finding for modellers is that one can no longer hope or pretend to simulate the full network behavior by assembling neurons with fixed intrinsic or context-invariant properties in a pure bottom-up approach. Models of the early visual system have to incorporate homeostasis rules acting across integration levels to account for the inverse covariation between input drive complexity and the

apparent linearity of the more elementary processing units. In a kind of reciprocal way, it remains nevertheless highly plausible that the characterization of psychic laws at a more holistic level, guiding the psychological principles of low-level (non attentive) perception, pleads for the existence of structuro-functional traces that can be retrieved by decoding and averaging synaptic echoes. The identification at more microscopic scales of the key non-linearities and the activity synergy requirements necessary for their expression should help us to understand, in a causal way, the **emergence** of novel binding principles at more mesoscopic levels, leading eventually to the abstract formalization of psychological foundations of perception.

Acknowledgments Work was supported by CNRS, The Paris-Saclay IDEX (NeuroSaclay and I-Code), the French National Research Agency (ANR: NatStats and Complex-V1) and the European Community (FET-Bio-I3 integrated programs (IP FP6: FACETS (015879), IP FP7: BRAINSCALES (269921); FET-Open (Brain-i-nets (243914); FET-Flagship: The Human Brain Project).

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Alonso JM (2002) Neural connections and receptive field properties in the primary visual cortex. *Neuroscientist* 8(5):443–456
- Bai EW, Cerone V, Regruto D (2007) Separable inputs for the identification of block-oriented non-linear systems. In: *Proceedings of 2007 American Control Conference*, New York, pp 1548–1553
- Barlow HB (1972) Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1:371–394
- Baudot P, Levy M, Marre O, Monier C, Pananceau M, Frégnac Y (2013) Animation of natural scene by virtual eye-movements evokes high precision and low noise in V1 neurons. *Front Neural Circ* 7:206–235
- Benucci A, Frazor RA, Carandini M (2007) Standing waves and traveling waves distinguish two circuits in visual cortex. *Neuron* 55(1):103–117
- Borg-Graham LJ, Monier C, Frégnac Y (1998) Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature* 393:369–373
- Bosking WH, Zhang Y, Schofield B, Fitzpatrick D (1997) Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J Neurosci* 1(6):2112–2127
- Bringuier V, Chavane F, Glaeser L, Frégnac Y (1999) Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science* 283:695–699
- Buzás P, Kovács K, Ferecskó AS, Budd JM, Eysel UT, Kisvárdy ZF (2006) Model-based analysis of excitatory lateral connections in the visual cortex. *J Comp Neurol* 499(6):861–881
- Callaway EM, Katz LC (1990) Emergence and refinement of clustered horizontal connections in cat striate cortex. *J Neurosci* 10(4):1134–1153

- Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA (2005) Do we know what the early visual system does? *J Neurosci* 25(46):10577–10597
- Chance FS, Nelson SB, Abbott LF (1999) Complex cells as cortically amplified simple cells. *Nat Neurosci* 2:277–282
- Chavane F, Sharon D, Jancke D, Marre O, Frégnac Y, Grinvald A (2011) Lateral spread of orientation selectivity in V1 is controlled by intracortical cooperativity. *Front Syst Neurosci* 5:4–24. doi:10.3389/fnsys.2011.00004
- Chen X, Han F, Poo M-M, Dan Y (2007) Excitatory and suppressive receptive field subunits in awake monkey primary visual cortex (V1). *Proc Natl Acad Sci USA* 104:19120–19125
- Douglas RJ, Martin KA (2004) Neuronal circuits of the neocortex. *Annu Rev Neurosci* 27:419–451
- Ferster D, Miller KD (2000) Neural mechanisms of orientation selectivity in the visual cortex. *Annu Rev Neurosci* 23:441–471
- Field DJ, Hayes A, Hess RF (1993) Contour integration by the human visual system: evidence for a local “association field”. *Vision Res* 33(2):173–193
- Fournier J, Monier C, Pananceau M, Frégnac Y (2011) Adaptation of the simple or complex nature of V1 receptive fields to visual statistics. *Nat Neurosci* 14(8):1053–1060
- Fournier Y, Monier C, Levy M, Marre O, Sari K, Kisvarday ZF, Frégnac Y (2014) Hidden complexity of synaptic receptive fields in cat primary visual cortex. *J Neurosci* 34(16):5515–5528
- Frégnac Y (2012) Reading out the synaptic echoes of low level perception in V1. *Lect Notes Comput Sci* 7583:486–495
- Frégnac Y, Rudolph M, Davison A, Destexhe A (2007) Complexity and level hierarchy in neural networks. In: Képès F (ed) *Biological networks. Complex systems and interdisciplinary science series*. World Scientific, Singapore, pp 291–340
- Frégnac Y, Bathellier B (2015) Cortical correlates of low-level perception: from neural circuits to percepts. *Neuron* 88:110–126
- Gérard-Mercier F, Carelli P, Pananceau M, Baudot P, Troncoso X, Frégnac Y (2014) A saccadic view of the “silent surround” of visual cortical receptive fields. *American Society for Neuroscience Abstracts*, Washington, DC
- Gerard-Mercier F, Pananceau M, Carelli P, Troncoso X, Frégnac Y (2016) Synaptic correlates of low-level perception in V1. *J Neurosci* (submitted)
- Gilbert CD, Li W (2012) Adult visual cortical plasticity. *Neuron* 75(2):250–264
- Gilbert CD, Wiesel TN (1983) Clustered intrinsic connections in cat visual cortex. *J Neurosci* 3(5):1116–1133
- Haider B, Krause MR, Duque A, Yu Y, Touryan J, Mazer JA, McCormick DA (2010) Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron* 65(1):107–121
- Henry GH (1977) Receptive field classes of cells in the striate cortex of the cat. *Brain Res* 133:1–28
- Huang X, Elyada YM, Bosking WH, Walker T, Fitzpatrick D (2014) Optogenetic assessment of horizontal interactions in primary visual cortex. *J Neurosci* 2(34):4976–4990 [Erratum in *J Neurosci* 34(26): 8930]
- Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J Physiol (London)* 160:106–154
- Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol (London)* 195:215–243
- Hubel DH, Wiesel TN (2005) *Brain and visual perception*. Oxford University Press, New York
- Jones HE, Grieve KL, Wang W, Sillitio AM (1980) Surround suppression in primate V1. *J Neurophysiol* 86:2011–2028
- Judge SJ, Wurtz RH, Richmond BJ (1980) Vision during saccadic eye movements. I. Visual interactions in striate cortex. *J Neurophysiol* 43(4):1133–1155
- Kasamatsu T, Miller R, Zhu Z, Chang M, Ishida Y (2010) Collinear facilitation is independent of receptive field expansion at low contrast. *Exp Brain Res* 201(3):453–465
- Kisvarday ZF, Toth E, Rausch M, Eysel UT (1997) Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. *Cereb Cortex* 7:605–618
- Li W, Piéch V, Gilbert CD (2006) Contour saliency in primary visual cortex. *Neuron* 50(6):951–962

- Li W, Piéch V, Gilbert CD (2008) Learning to link visual contours. *Neuron* 57(3):442–451
- Martin KAC (2014) Obituary David Hubel (1926–2013). *Curr Biol* 24(1):R4–R7
- Martin KAC, Roth S, Rusch ES (2014) Superficial layer pyramidal cells communicate heterogeneously between multiple functional domains of cat primary visual cortex. *Nat Commun* 5:5252–5265. doi:10.1038/ncomms6252
- Mizobe K, Polat U, Pettet MW (2001) Facilitation and suppression of single striate-cell activity by spatially discrete pattern stimuli presented beyond the receptive field. *Vis Neurosci* 18(3):377–391
- Monier C, Chavane F, Baudot P, Graham LJ, Frégnac Y (2003) Orientation and direction selectivity of synaptic activity in visual cortical neurons: a diversity of combinations produces spike tuning. *Neuron* 37:663–680
- Monier C, Fournier J, Frégnac Y (2008) In vitro and in vivo measures of evoked excitatory and inhibitory conductance dynamics in sensory cortices. *J Neurosci Methods* 169:323–365
- Nauhaus I, Busse L, Carandini M, Ringach DL (2009) Stimulus contrast modulates functional connectivity in visual cortex. *Nat Neurosci* 12:70–76
- Olshausen BA (2014) 20 years of learning about vision: questions answered, questions unanswered, and questions not yet asked. In: Bower JM (ed) 20 years of computational neuroscience. Springer, New York. ISBN 978-1-4614-1424-7
- Polat U, Sagi D (1985) Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Res* 33(7):993–999
- Polat U, Sagi D (1993) Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Res* 33:993–999
- Priebe NJ, Ferster D (2012) Mechanisms of neuronal computation in mammalian visual cortex. *Neuron* 75(2):194–208
- Priebe NJ, Mechler F, Carandini M, Ferster D (2004) The contribution of spike threshold to the dichotomy of cortical simple and complex cells. *Nat Neurosci* 7(10):1113–1122
- Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2005) Spatiotemporal elements of macaque v1 receptive fields. *Neuron* 46:945–956
- Sillito AM, Jones HE (1996) Context-dependent interactions and visual processing in V1. *J Physiol (Paris)* 90(3–4):205–209
- Sillito AM, Grieve KL, Jones HE, Davis J (1995) Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* 378(6556):492–496
- Skottun BC, De Valois RL, Grosof DH, Movshon JA, Albrecht DG, Bonds AB (1991) Classifying simple and complex cells on the basis of response modulation. *Vision Res* 31:1079–1086
- Tan AYY, Chen Y, Scholl B, Seidemann E, Priebe NJ (2014) Sensory stimulation shifts cortex from synchronous to asynchronous states. *Nature* 509(8):226–230
- Tao L, Shelley M, McLaughlin D, Shapley R (2004) An egalitarian network model for the emergence of simple and complex cells in visual cortex. *Proc Natl Acad Sci USA* 101:366–371
- Touryan J, Lau B, Dan Y (2002) Isolation of relevant visual features from random stimuli for cortical complex cells. *J Neurosci* 22(24):10811–10818
- Touryan J, Felsen G, Dan Y (2005) Spatial structure of complex cell receptive fields measured with natural images. *Neuron* 45:781–791
- Troncoso X, Pananceau M, Lebec B, Desbois C, Gerard-Mercier F, Frégnac Y (2015) Spatial synergy and temporal coherence requirements for binding the feedforward and horizontal waves in V1. American Society for Neuroscience Abstracts, Chicago, IL
- Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287(5456):1273–1276
- Wagemans J, Edler JH, Kubovy M, Palmer SE, Peterson MA, Singh M, von der Heydt R (2012) A century of Gestalt psychology in visual perception. I. Perceptual grouping and figure-ground organization. *Psychol Bull* 138(6):1172–1217
- Yeh CI, Xing D, Williams PE, Shapley RM (2009) Stimulus ensemble and cortical layer determine V1 spatial receptive fields. *Proc Natl Acad Sci USA* 106:14652–14657

Grid Cells and Spatial Maps in Entorhinal Cortex and Hippocampus

Tor Stensola and Edvard I. Moser

Abstract The cortical circuit for spatial representation has multiple functionally distinct components, each dedicated to a highly specific aspect of spatial processing. The circuit includes place cells in the hippocampus as well as grid cells, head direction cells and border cells in the medial entorhinal cortex. In this review we discuss the functional organization of the hippocampal-entorhinal space circuit. We shall review data suggesting that the circuit of grid cells has a modular organization and we will discuss principles by which individual modules of grid cells interact with geometric features of the external environment. We shall argue that the modular organization of the grid-cell system may be instrumental in memory orthogonalization in place cells in the hippocampus. Taken together, these examples illustrate a brain system that performs computations at the highest level, yet remains one of the cortical circuits with the best readout for experimental analysis and intervention.

Place Cells and Grid Cells

An entirely new branch of neuroscience opened with the discovery of hippocampal place cells, i.e., cells that fire specifically when animals are in certain locations (O'Keefe and Dostrovsky 1971; O'Keefe and Nadel 1978; Fig. 1). Different place cells were found to fire in different locations of the environment (O'Keefe 1976), such that, for any given ensemble of place cells, the animal's location could be decoded from the pattern of activity among those cells (O'Keefe and Nadel 1978; Wilson and McNaughton 1993). With these insights, place cells became a strong

T. Stensola

Kavli Institute for Systems Neuroscience and Centre for Neural Computation, Norwegian University of Science and Technology, Olav Kyrres gate 9, 7491 Trondheim, Norway

Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Lisbon 1400-038, Portugal

E.I. Moser (✉)

Kavli Institute for Systems Neuroscience and Centre for Neural Computation, Norwegian University of Science and Technology, Olav Kyrres gate 9, 7491 Trondheim, Norway
e-mail: edvard.moser@ntnu.no

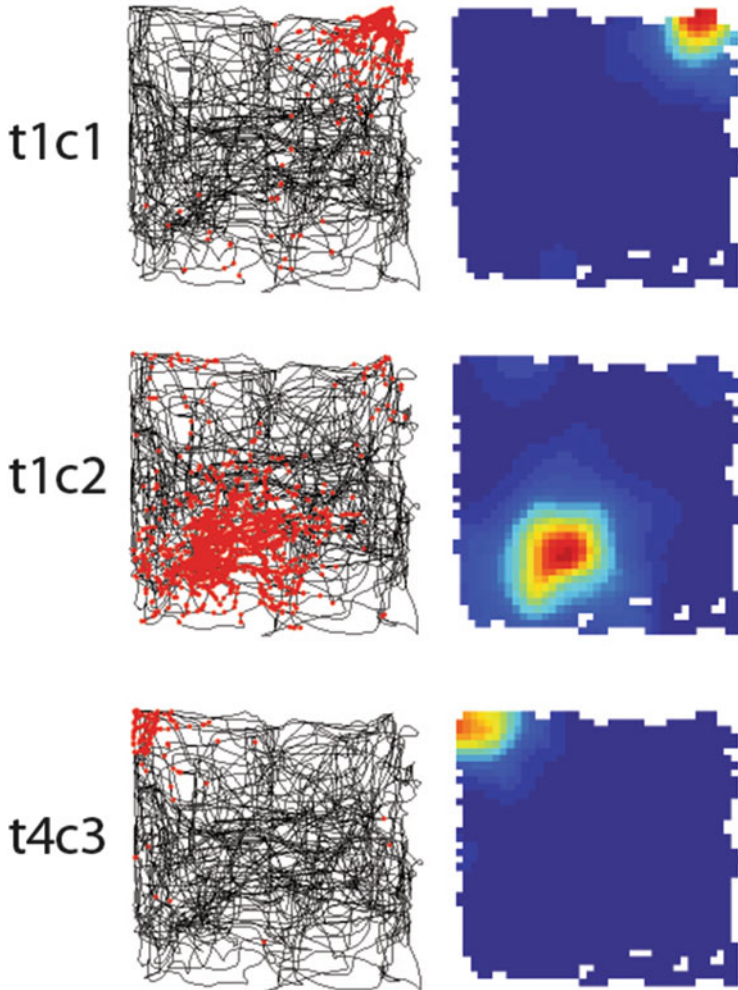


Fig. 1 Place cells recorded in hippocampal subarea CA3. Bird's eye view of firing locations of three place cells, with firing locations shown as *red dots* on the path of the rat (*black*). *t* indicates tetrode number, *c* cell number. Cells were recorded simultaneously. *Right*: pseudo-color activity maps of the cells to the left. *Red* is high firing rate, and *blue* is no firing. Reproduced with permission from Fyhn et al. (2007)

candidate for the neural implementation of Tolmanian cognitive maps, maps that animals use to guide their navigation in the environment (Tolman 1948; O'Keefe and Nadel 1978).

In trying to understand which incoming signals could take part in generating location-specific responses in place cells, both experimental and theoretical suggestions have been presented. An important clue was the experimental

demonstration that place cells in CA1 could sustain place characteristics after ablation of all input from CA3 (Brun et al. 2002). This observation suggested that place responses in CA1 originated from an alternative source of excitatory input to CA1: the medial entorhinal cortex (MEC). In pursuing this possibility, we observed that neurons in MEC were also spatially selective (Fyhn et al. 2004; see also Hargreaves et al. 2005), although MEC neurons typically had several firing fields in environments where place cells had only a single field. It turned out that the firing fields of the spatial cells in MEC formed a near-perfect hexagonal grid tessellating the entire space available to the animal (Hafting et al. 2005; Fig. 2). Each grid cell had a slightly different set of x, y -coordinates in the environment, so that the entire environment could be covered collectively by a small number of grid cells. Dorsally in MEC, grid patterns typically had small fields packed densely together. At more ventral MEC locations, with increasing distance from the dorsal MEC border, the scale of the grid pattern expanded (Fyhn et al. 2004; Hafting et al. 2005; Brun et al. 2008; Fig. 2). Several computational models (O'Keefe and Burgess 2005;

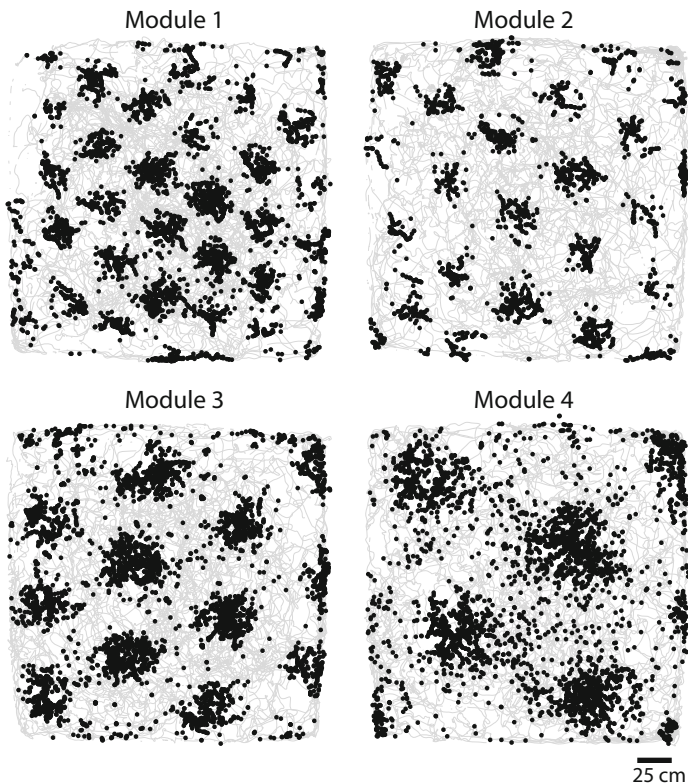


Fig. 2 Grid cell firing patterns; bird's eye view. Action potentials (*black*) superimposed on the movement path (*gray*) reveal a periodic spatial activity pattern. Shown are grid patterns of four distinct scales recorded within the same animal. Reproduced with permission from Stensola et al. (2012)

Fuhs and Touretzky 2006; McNaughton et al. 2006; Burak and Fiete 2009; Burgess et al. 2007) and multiple lines of experimental evidence (Brun et al. 2002; Van Cauter et al. 2008; Zhang et al. 2013) soon pointed to grid cells as prime candidates in conferring spatial selectivity to place cells in downstream hippocampus.

Models that describe possible grid-to-place transforms are dependent on how the grid map is organized at several functional levels. Grid spacing is organized topographically along the dorsoventral axis of MEC, with average grid spacing increasing from dorsal to ventral (Fyhn et al. 2004; Hafting et al. 2005; Brun et al. 2008). Despite initial reports based on low cell numbers (Barry et al. 2007), it remained unclear after the first studies whether grid scale distributed within animals as a scale-continuum or instead progressed in steps. To answer this question, it was essential to record large numbers of grid cells over considerable dorsoventral distances within animals, so as to sample a sufficient range of grid spacing. It was necessary to record with minimal discontinuity in the tissue so that steps in spacing could be discerned reliably from discontinuities in sampling of a smooth topography.

In the first reports of grid cells (Hafting et al. 2005; Fyhn et al. 2007), co-localized cells always had a similar grid orientation (orientation of grid axes), suggesting there was only one shared orientation in the entire circuit. Later work has shown that multiple orientation configurations may be present in the same animal (Krupic et al. 2012; Stensola et al. 2012). The existence of multiple orientation configurations across multiple levels of grid scale highlights a basic question: is the grid map composed of smaller sub-maps or does it act as one coherent representation of space, but with variable geometric features such as spacing and orientation? A grid map with independently functioning sub-maps may produce unique population-pattern combinations for every environment, resulting in unique input patterns to place cells and, in turn, unique hippocampal output (Fyhn et al. 2007). A major objective, based on this possibility, has therefore been to determine if grid cells within the same grid circuit perform separate operations on the same inputs. The next section will address the possibility of a modular functional organization of the grid-cell circuit.

Discretization of the Entorhinal Grid Map

Locally, grid cells behave as a coherent ensemble (Fyhn et al. 2007), but it was unknown from the first reports if the entire grid map functioned as a coordinated whole or if it was fractioned into sub-units that displayed a capacity for independent function. By combining novel and established experimental approaches, we were able to record an unprecedented number of grid cells—up to 186 cells from the same animal—which finally allowed us to determine that the grid map is a conglomerate of sub-maps or modules (Stensola et al. 2012).

The new recordings showed, within animals, that the gradient in grid scale (grid spacing) along the dorsoventral axis of MEC progressed in clear steps rather than as

a continuum. All cells within a module shared the same grid spacing, and modules of increasing scale became more abundant as the tetrodes were turned to more ventral MEC locations. Cells that shared the same grid spacing within animals also had a common grid orientation, defined as the orientation of the grid axes relative to the local boundaries of the environment. Most grid cells also demonstrated small but consistent deviations from perfect hexagonal symmetry, expressed by the fact that the inner ring of fields in the grid pattern formed an ellipse rather than a circle. These deformations were consistent across cells in the same grid module (Stensola et al. 2012). No modular organization was apparent within the population of head direction cells in the MEC (Giocomo et al. 2014).

Modular organization was also expressed in the temporal modulation of spike activity. Grid cells are tuned to the ongoing population activity, manifested as oscillations in the local field potential (Hafting et al. 2008; Jeewajee et al. 2008). Several models implicate theta rhythms in the generation of the grid pattern (Burgess et al. 2007; for review, see Moser et al. 2014). Previous work had shown that cells at ventral locations of the dorsoventral MEC axis oscillated with a slower beat frequency than dorsal cells, and it was suggested that this gradient arose from gradients in the expression of specific ion channels (Giocomo et al. 2007; Giocomo and Hasselmo 2008; Garden et al. 2008; Jeewajee et al. 2008). We found that grid cells in geometrically defined modules were modulated by the same theta frequency. On average, modules with greater grid spacing had lower theta frequencies, but within animals, modules were not strictly confined to this trend.

The consistency of geometric features within but not across modules made it possible to define module membership for all cells with an automated multidimensional clustering approach (K-means clustering). After defining the modules, we could turn to the question of how modules were distributed in the MEC tissue. Several signs of anatomical clustering existed within the entorhinal system (Ikeda et al. 1989; Solodkin and Van Hoesen 1996; Burgalossi et al. 2011), pointing to possible anatomical substrates for the functional clustering. Individual modules occupied extensive portions of MEC. We found that, on average, a module spanned >1 mm of the dorsoventral MEC axis. There was extensive module overlap in the intermediate-to-ventral parts of MEC such that, at any MEC location, cells from several modules could be present. Grid modules were found to cut across cell layers; cells that were part of one module were found in several layers. In contrast to the organization along the dorsoventral axis, there was no discernable topography along the mediolateral axis. Instead, modules extended across large mediolateral distances (~1 mm, which was the limit of our recording arrays), suggesting modules distributed as mediolateral bands along the dorsoventral axis. Based on this knowledge, combined with the distribution of modules along the dorsoventral axis, we could estimate the number of distinct modules within animals to be in the upper single-digit range. This anatomical distribution of modules does not match any known anatomical clustering in the entorhinal cytoarchitecture.

With previous reports having suggested a set relationship between scale steps (Barry et al. 2007), we next quantified the relationship between module scales

within and across animals. Within animals, there was considerable variation in the relationship between one module scale and the next, suggesting that scale is set independently for each module and animal. However, when we pooled the scale progression across animals, a pattern was revealed. On average, modules increased by a fixed scale ratio, as a geometric progression. The mean ratio was 1.42, very close to $\sqrt{2}$. This relationship pointed to genetic circuit-mechanisms as contributors to grid scale, yet the geometric individuality of the modules suggested that modules exhibited a substantial level of autonomy.

Finally, in a separate set of experiments, we tested if grid modules were also functionally independent. Grid cells are known to rescale along with environmental compression (Barry et al. 2007; Solstad et al. 2008). We found that, when animals were exposed to a relocation of one of the walls in the environment, modules rescaled along the compression, but to varying degrees (Stensola et al. 2012). Cells within a module behaved coherently, whereas individual modules could rescale to completely different extents within animals. This finding provided the first proof-of-principle for independent function within sub-populations in the grid map.

Combinatorics in Grid Cells and Remapping in Place Cells

In a landmark study of place cells, Muller and Kubie (1987) described a phenomenon that had great implications for our understanding of the relationship between the spatial map in hippocampus and its role in memory formation. For one, they demonstrated, in agreement with earlier work (O'Keefe and Conway 1978), that place cells were under the control of sensory cues in the environment, as rotation of a cue resulted in consistent rotation of the place fields. More importantly, they showed that, if two recording environments differed beyond a certain magnitude, the activity of the recorded cells changed drastically between the environments. Among the cells that were active in the first environment and remained active in the second, the firing locations were completely reorganized in space. Further, a large portion of cells that were active in one environment became silent in the next. Other cells were active only in the second environment. This functional reorganization was termed 'remapping' and represented an orthogonalization in the population encoding between the distinct environments.

Grid modularity appears to offer very favorable conditions for hippocampal remapping (Fig. 3). Maps from different grid modules could reorganize to yield completely novel downstream population inputs and, therefore, new hippocampal place maps. Early work showed that grid cells realigned with the environment when remapping took place in simultaneously recorded hippocampal place cells (Fyhn et al. 2007). The realignment involved a shift in grid phase and a reorientation of the grid pattern relative to the geometry of the environment. The realignment was coherent for all grid cells recorded, so that spatial relationships between the grid cells remained. This observation does not preclude independent realignment of distinct modules, however, because all of the grid cells in the early work were

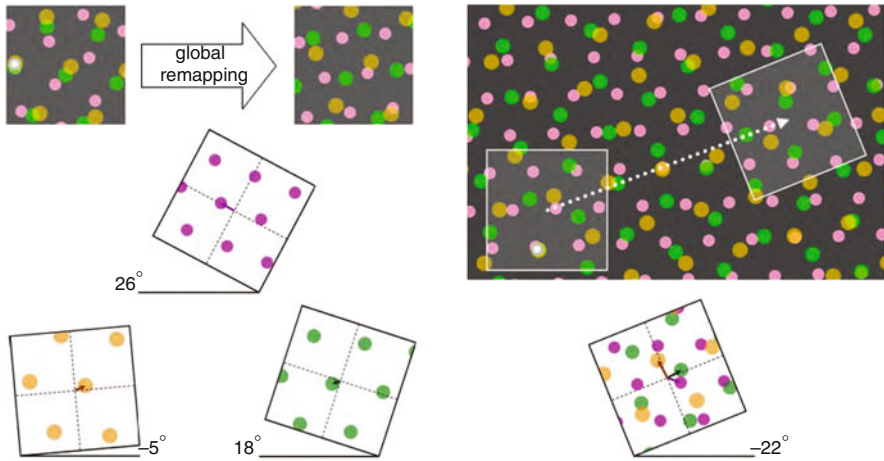


Fig. 3 Two proposed mechanisms that may underlie hippocampal remapping based on grid inputs. *Left*: several independent grid maps, each with a different *color*, realign independently (*bottom*) and cause unique combinatorial population patterns in the hippocampus (*top*). *Right*: the grid map is coherent across scales, and remapping occurs from a shift in spatial phase space. Reproduced with permission from Fyhn et al. (2007)

recorded at the dorsal end of the MEC and all had a relatively similar grid scale, i.e., most of the cells may have belonged to the same module.

If grid modules are the main source of hippocampal remapping, the level of independence between grid modules will affect remapping-based mnemonic capacity. But how independent are the grid modules? Grid modules have several geometric traits that suggest autonomy (Stensola et al. 2012). Grid spacing relationships varied across animals, and grid orientation could be completely offset between modules. Grid modules also differed in the amount and directionality of pattern deformation, and deformation, scale and orientation changed independently across modules when the animal was exposed to a novel room (Fig. 4). These observations are entirely consistent with an attractor mechanism for grid formation. In attractor models of grid cells, a grid network can only support a circuit in which all cells share the same geometry (McNaughton et al. 2006; Burak and Fiete 2009; Moser et al. 2014).

A surprising observation, however, was that modules typically assumed one of only four distinct orientation configurations relative to the environment (Stensola et al. 2015). This constraint on orientation may seem highly disadvantageous for generating maximally distinct hippocampal inputs. However, it has been shown theoretically that remapping based on grid modules is much more sensitive to the spatial phase offset between the modules than the relative orientation and spacing (Monaco and Abbott 2011; Fig. 5). Varying grid orientation caused less reorganization in hippocampus compared to varying phase.

The differences in rescaling across grid modules may shed light on the mechanisms underlying rescaling of hippocampal place fields after changes in the

Rat:14147

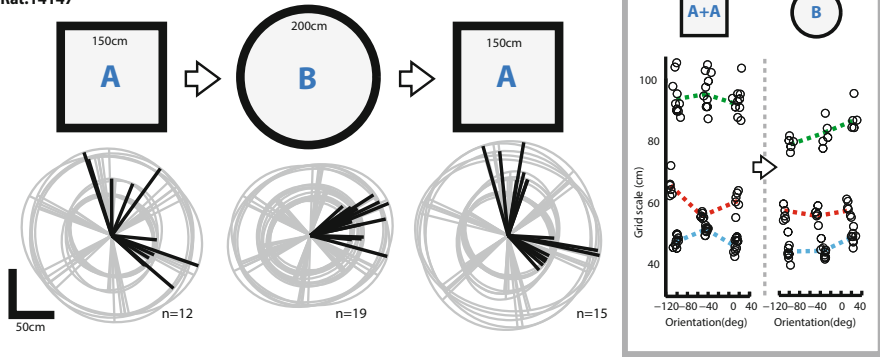


Fig. 4 Modules realigned when animals were tested in a novel box in a novel room. Grid scale, orientation and ellipse directions all changed independently between modules, strongly suggesting independent operations. The *left panel* shows grid axes and ellipse (gray lines) and ellipse tilt (black line) from all cells in one animal in square and circular environments. Note the independent changes in ellipse tilt. The *figure on the right* shows data from all three grid axes in the square and the circle. Reproduced, with permission, from Stensola et al. (2012)

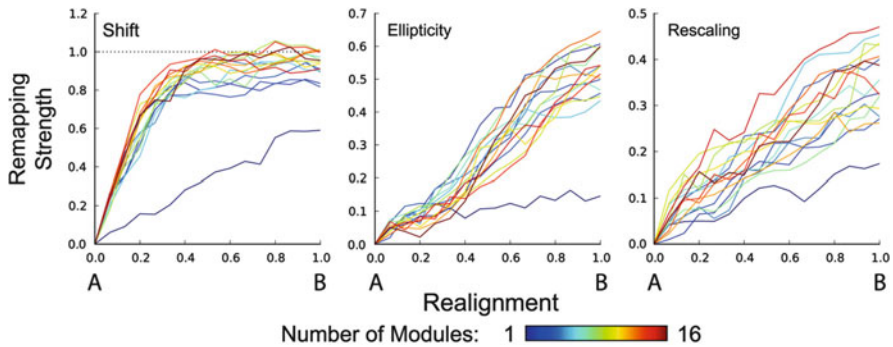


Fig. 5 Efficacy of reorganizing different parameters of grid geometry between modules. The strongest remapping occurred from phase shifts, while other features (changes in elliptic deformation or scale) were less effective. A and B denote the two distinct environments. Reproduced, with permission, from Monaco and Abbott (2011)

geometry of the environment (O'Keefe and Burgess 1996). O'Keefe and Burgess recorded place cells in a rectangular environment that could be extended or compressed in any of the four cardinal directions. When the recording box was extended or compressed, place fields followed the change in environmental geometry. Some cells were anchored to one wall or a set of walls so that their firing fields moved along with the extension. Other cells were anchored to the external room instead of the box, and yet others distended the place field along the box or even split the field in two parts. This behavior suggested an underlying input pattern with a distinct geometric relationship to the walls of the recording box or the room.

Based on their observations, the authors proposed a model in which spatial modulation arose from the sum of multiple Gaussian activity bands offset from the environmental boundaries at different distances (O'Keefe and Burgess 1996). This idea was later developed into the boundary-vector model of place cells (Hartley et al. 2000; Barry et al. 2006). Although boundary-selective cells exist in MEC and do project to hippocampal place cells (Zhang et al. 2013), this study is also intuitively in line with expectations from the observations of grid rescaling. Because of rescaling, place fields can receive input that is topologically identical to the original map, only distended or compressed, likely resulting in distended or compressed place fields. If a place cell receives input from two modules, and these modules differ greatly in rescaling, it seems reasonable to assume that their contribution is split into two fields under some circumstances.

Topography in Parahippocampal Systems

In sensory and motor cortices, there is often a neat correspondence between relationships in the external world and their internal representations in the brain, that is to say, continuous or discrete variables in the external world are mapped topographically into the cortical sheet (Rasmussen and Penfield 1947). Topography often represents a 'where' component onto which information about stimulus quality can be superimposed (Tolman 1948; Montagnini and Treves 2003; Kaas 2012). Similar correspondences are not present in the entorhinal-hippocampal system. Neither place cells nor grid cells display topographic representation of spatial location, at least not at the macroscopic level (O'Keefe 1976; Redish et al. 2001; Hafting et al. 2005). In the hippocampus, with two-photon imaging, the activity of an entire ensemble of hippocampal place cells could be imaged simultaneously while a mouse navigated within a virtual environment (Dombeck et al. 2010; Harvey et al. 2009). The place cells developed clear place fields, suggesting the task was not too alienating for the spatial representation system. Further, there was no statistical relationship between the location of place cells in the hippocampal cellular sheet ($>35 \mu\text{m}$ apart), and the locations of their place fields in virtual space. Cell pairs $<35 \mu\text{m}$ apart, displayed a significant correlation but not separable from correlations from common neuropil or activity bleed-over in the imaging technique (Dombeck et al. 2010). Imaging of grid cells using similar methods has recently confirmed that grid phase is organized non-topographically at the macroscopic level, but larger-than-expected correlations were reported for cells that were nearest neighbors (Heys et al. 2014).

The low level of topography in the adult spatial representation system does not exclude the existence of topography at earlier developmental stages. Mechanisms by which topography is present during development as a teaching signal to set up appropriate circuitry for grid function, only to disappear in the adult brain, have been proposed for place cells (Samsonovich and McNaughton 1997) as well as grid cells (McNaughton et al. 2006). It is worth noting that both hippocampus and

entorhinal cortex are evolutionarily ‘old,’ such that the orderly topography seen in typical low level cortex only likely arose after these structures were past their phylogenetic window of opportunity (Kaas 2012). The olfactory piriform cortex is another ancient cortical structure that does not show topographical organization, even though continuity in stimulus dimensions exists and similar teaching inputs may have been present.

The apparent lack of topographical mapping of firing locations contrasts with the progressive increase in average scale of place cells (Jung et al. 1994; Kjelstrup et al. 2008) and grid cells (Fyhn et al. 2004; Hafting et al. 2005; Brun et al. 2008) along the dorsoventral axis of the hippocampus and the MEC, respectively. What are the functional consequences of this scale expansion? There is an extensive literature on the distinct features of dorsal and ventral portions of the hippocampus. Lesions at different dorsoventral portions produce markedly different behavioral deficits (Nadel 1968; Moser et al. 1993). Lesions of a small portion of the dorsal pole impair spatial memory efficiently, whereas similar portions of the ventral pole do not (Moser et al. 1993, 1995). Stress responses and emotional behavior are affected by lesions to ventral but not dorsal portions of hippocampus (Henke 1990; Kjelstrup et al. 2002). Connectivity to and from these portions of hippocampus is distinct (Witter et al. 1989; Dolorfo and Amaral 1998). There is also a growing body of literature in spatial cognition in humans suggesting functional polarization along the human equivalent of the dorsoventral axis (Fanselow and Dong 2010; Poppenk et al. 2013). In particular, activity in the human equivalent of the ventral hippocampus is associated with coarse global spatial representations and route planning and execution, whereas the dorsal equivalent is associated with fine-grained local representations and navigation strategies, such as number of turns on a route (Evensmoen et al. 2013).

The neural codes along the dorsoventral axis of the parahippocampal spatial system may very well reflect an axis of generalization. With increased scale of spatial fields in the hippocampus and the MEC, the larger fields do not denote spatial location with equal demarcation, so spatial resolution is diminished. Another consequence is that for these ventral codes, at any particular point in space, a greater portion of cells will be active. This increase in representational density may confer better robustness to noise: the more cells that can take part in a ‘majority’ vote, the better the vote will be statistically, despite poorer spatial resolution. Exactly the same argument can be made for the representation of head direction, whose resolution also decreases from dorsal to ventral MEC (Giocomo et al. 2014; Fig. 6). Alternatively, ventral cells (both grid, place and head direction cells) code a larger portion of the environment at any moment, so that the population code at any location is more generalized. This may be beneficial for associating content into current spatial contexts. The ventral hippocampus is more associated with stress and fear responses and has stronger connections with the amygdala (Moser and Moser 1998). For embedding fear memories into spatial context, it may be advantageous to impose a higher level of generalization.

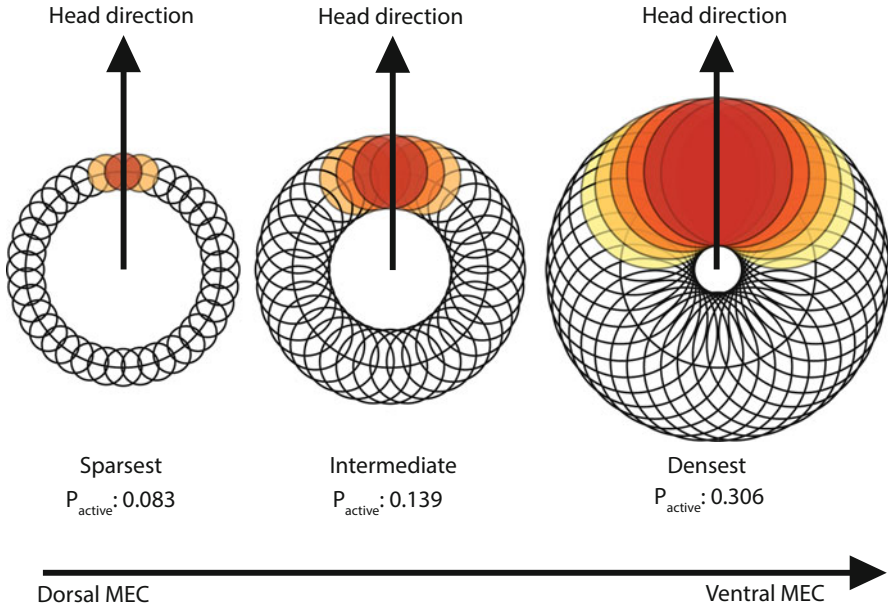


Fig. 6 Head direction representational density increases along the dorsoventral axis in MEC layer 3. Each doughnut represents a head direction cell population, and each cell is represented as a *circle* on the doughnut. The location and size of the *circle* represent preferred head direction and tuning specificity, respectively. Given populations of equal size (same number of rings; dorsal to ventral as *left to right*), and the same directional input, ventral populations will have a larger proportion (P) of its cell population be active to any input compared to more dorsal populations due to broader tuning (*color gradient* shows each cell's activity level; *red* is maximum)

Mechanisms of Grid Spacing

Our studies have shown that grid spacing increases in steps along the dorsoventral axis of MEC. The factors that determine topographical grid spacing are currently unknown. When all module pair ratios were pooled across animals, a consistent average scale ratio was revealed. This consistency across animals implies a genetic component in determining grid spacing. Gradients of specific ion channels, such as the hyperpolarization-activated cyclic nucleotide-gated (HCN) channels, exist in entorhinal cortex and have been suggested to account for the topography of grid scale (Giocomo et al. 2007, Giocomo and Hasselmo 2009; Garden et al. 2008). However, such channels, when genetically knocked out, did not remove grid scaling along the dorsoventral axis but instead changed the baseline spacing (Giocomo et al. 2011). Other channel gradients may contribute to scaling, such as potassium channels (Garden et al. 2008). If scale is determined in part from channel gradients, or indeed any genetic expression pattern, it seems likely the gradient will provide a smooth topography of any conferred scale parameter, instead of a

modular organization. How then could modular grid scale result from a smooth underlying gradient?

One possible scenario is that module grid scale is determined by network dynamics acting on a graded underlying scale parameter. Attractor models of grid cells predict that all cells in a circuit must have the same grid spacing (as well as orientation and pattern deformation) to generate a stable grid pattern (Welinder et al. 2008). Within a grid network determined by attractor dynamics, there will likely be some tolerance to small variations in the scale-parameter distribution across cells, so that when the network is initiated, the effects of population dynamics dominate individual cells enough to coordinate all cells into a common pattern, cancelling out individual variation. In a sense, this ‘spatial synchronization’ acts similarly to synchronization in the temporal domain; originally observed by Huygens in 1665, coupled oscillators settle on a mean frequency that entrains all the individual oscillators, even in the presence of relatively large variations in individual frequencies. But what would happen if the scale parameter distribution exhibits too large spread? The variation may become too large to entrain all units into one coherent pattern, and the pattern may fraction into sub-ensembles that each center on a mean frequency that the ensemble can sustain. This way, by having a network self-organize from a very wide, continuous scale parameter distribution, such as channel expressions along an axis in MEC, several local modules of internal spatial consistency could arise from the unstable global pattern.

We observed convincing signs of independence between modules within animals, in terms of pattern geometry and rescaling responses. To incorporate this finding into the suggested mechanism above, one can suppose that, during development, learning strengthens connections within spatially synchronized ensembles but weakens connections between spatially desynchronized cells. In agreement with this possibility, grid cell pairs with similar spatial phase show stronger functional connectivity than pairs with dissimilar phase (Dunn et al. 2015). If two cells have a similar spatial phase, their coordinated firing in space will cause coordinated firing in time, a prerequisite for many forms of long-term potentiation (LTP; Bi and Poo 1998). Enhancement of connections between grid cells with a similar phase would lead to the development of functional ensembles intermingled in the same tissue, with strong inter-ensemble connectivity and weak cross-ensemble connectivity, in effect decoupling the ensembles functionally. A testable prediction from this idea is that very young animals, which have yet to achieve complete module decoupling, will display grid cells with poor spatial regularity because the network cannot sustain a coherent grid pattern based on cross-ensemble interactions. As the animal explores more space, decoupling will at some point become complete enough for cells to self-organize into modules with coherent and regular grid patterns. Such a transition may be rapid, as it may involve a ‘tipping point’ after which network dynamics kick in to entrain the ensemble. In two studies that characterized grid cells in early development in rats, grid patterns were indeed not very regular initially (Langston et al. 2010; Wills et al. 2010). Only at the age of about 4 weeks, 1–2 weeks after the beginning of outbound exploration, did regular

grid firing occur. The transition to this state had a rapid onset, in line with the above proposal.

Conversely, if the scale parameter is associated with temporal characteristics such as intrinsic resonance frequency, as suggested in several models (Burgess et al. 2007) and by experimental findings (Giocomo et al. 2007, 2011; Jeewajee et al. 2008), synchronization in the temporal domain during development could result in similar module fractionation and synaptic modification to cause *temporally* consistent ensembles. If the scale parameter is associated with temporal frequency, these temporally synchronized ensembles would also become spatially synchronized. By this mechanism, grid modules could develop to mature, functionally decoupled modules at least in part before the animal ever explores space. In line with this is our finding that modules are temporally consistent.

Shearing-Induced Asymmetries in Grid Cells

There are no hexagonal features in the environment that correspond to the grid pattern. Grid patterns are instead believed to arise from local network dynamics, with self-motion input as a major update mechanism (Fuhs and Touretzky 2006; McNaughton et al. 2006; Welinder et al. 2008; Couey et al. 2013). However, for the grid pattern to be useful in allocentric representations, it must anchor to the external environment. Several features of the pattern could be involved in this anchoring process, including spatial phase (offset in the x, y -plane), grid spacing and grid orientation (alignment between grid axes and axes of the environment). We demonstrated earlier that grid orientation can assume distinct orientations across and within animals (Stensola et al. 2012), but it was unknown whether there was any orderly relationship between grid alignment and specific features of the environment.

In a recent study (Stensola et al. 2015), we compared grid orientation from large data sets recorded in two distinct square environments, enabling rigorous analyses of grid alignment. Grid orientation did not distribute randomly across animals. Instead, there was a strong tendency for grid axes to align to the cardinal axes of the environment, defined by the walls of the recording enclosure. In one environment, we observed clustering around one wall axis only, whereas in the other environment grid orientation distributed around both cardinal axes. The strong bias towards the box axes suggested the box geometry itself acted as the grid anchor, and not salient extra-environmental visual cues, which were deliberately abundant.

Rather than aligning parallel to the box axes, cells were consistently offset from these axes by a small amount in all environments. In one environment, this offset was to either side of one cardinal axis. In the second environment, cells were also offset from parallel, but with reference to both cardinal axes. The rotation was identical across the two environments; cells were systematically offset from parallel by 7.5° , with a standard deviation of $2-3^\circ$, yielding four general alignment configurations for square environments. The observed distributions were not a result of

pooling across cells from different modules, as individual grid modules expressed the same absolute offset configurations, i.e., 7.5° .

What could be the function of the consistent offset of the grid axes? We noted that a triangular pattern within a square is maximally asymmetric at 7.5° rotation in relation to the axes of symmetry in the square, the same as the offset observed in the data. The environmental axes are primarily available to the animal in the form of borders imposed by the walls of the environment. Because border segments have been implicated in spatial coding (O'Keefe and Burgess 1996; Hartley et al. 2000; Barry et al. 2006) and because MEC contains cells that encode these specifically (Solstad et al. 2008; Savelli et al. 2008), we hypothesized that one function performed by grid alignment is to create maximally distinct population codes along border segments of the environment. This may be critical for encoding environments in which sensory input is ambiguous. Grid cells are thought to perform path integration (dead-reckoning from integration of distance and angle over time) based on self-motion cues. Without occasional sensory input, however, errors will accumulate until the representation becomes entirely unreliable. Sensory cues affect grid cells (Hafting et al. 2005; Savelli et al. 2008) and are thought to provide update signals that recalibrate path integration and reset accumulated errors. The symmetry and geometric ambiguity of the square recording environment may render such sensory cues less useful because multiple locations in the environment may produce similar update signals at different absolute locations. Therefore, error may be minimized by orientation solutions that maximize the distinctness of population representations at ambiguous locations.

Closer inspection showed that the angular offset of the grid axes differed between grid axes and depended on the angular distance from any of the walls of the square environment (Stensola et al. 2015). The further away a grid axis was from any of the walls, the smaller was the angular offset. The differential offset gave rise to an elliptic deformation of the circle described by the inner six fields of the grid pattern. The size and orientation of this elliptic deformation was not randomly distributed. In particular, the angular difference between the ellipse orientation of modules was clustered around 90 or 0° (Stensola et al. 2012). Because of this apparent link to the square geometry of the box geometry, we were inclined to investigate any possible links between elliptification of the grid and its offset. Ellipse orientation correlated strongly with angular offset, leading us to hypothesize that grid deformation and offset were the result of a common underlying process.

In continuum mechanics, simple shearing is a geometric transform that displaces points on a plane along a shear axis. Any point is displaced by an amount directly proportional to its Euclidian distance to that shear axis. The effect of this transformation on points that lie on a circle is that the circle becomes elliptic. Further, any axis on this circle will display non-coaxial rotation, the magnitude of which is directly proportional to the angular distance from the shear axis. To determine whether shearing could account jointly for the elliptic deformation and the angular offset of the grid, we applied shearing transformations to all grid patterns, with either of the cardinal box-wall axes as the shear axis (Stensola et al. 2015). Each

grid was sheared along each wall axis until it was minimally deformed, that is, least elliptical. We then determined how much the transform managed to reduce deformation, and how much the rotational offset was changed. We performed separate analyses for differently sized recording environments.

In the 1.5-m box, simple shearing removed most of the deformation (ellipticity was reduced from 1.17 to 1.06). It further completely removed the bimodal 7.5° offset peaks. The offset distribution became unimodal, with a peak centered close to 0° (parallel to one of the wall axes). This robust explanatory power of *simple* shearing implies that the grid pattern is globally anchored to one set of features such as a wall or a corner. We hypothesized that shearing develops with experience. In a smaller data set taken from a previous study (Fyhn et al. 2007), offset was indeed significantly closer to parallel in novel environment exposures than in familiar ones.

The 2.2-m box had more than twice the area of the 1.5-m box. Maintaining a coherent grid pattern may be sensitive to excessive distances between environmental features. If grid anchoring is globally set by a single feature (e.g., a border or wall), as suggested above, the integrity of the grid pattern may suffer at distances far from such anchoring points. We have shown previously that grid patterns fragment into local patterns in complex environments (Derdikman et al. 2009). We reasoned that, as the environment becomes larger, the grid pattern will benefit from stabilization by multiple anchoring points. In sufficiently large environments, spatial representation might break into locally anchored patterns that merge in the interior of the environment.

We applied simple shearing transforms to all grids from the 2.2-m box, exactly as with the smaller box. Minimizing deformation reduced ellipticity to the same extent, but the rotational offset was only partially removed. To test whether shearing occurred simultaneously from both wall axes, we determined for each cell the minimal deformation possible with a two-axis shearing transform. We could detect exactly one such minimum for every cell, suggesting it was one-to-one in the domain we were exploring. We then, as above, analyzed the impact on rotational offsets. The two-axis transform completely removed the offset peaks in the 2.2-m box, suggesting that the grid pattern had been sheared from two distinct anchoring sites.

A few modules did not display the common 7.5° offset and were not amenable to offset reduction through shearing. These modules nonetheless had 7.5° offsets locally in particular areas of the box. Such local offsets might not be detectable in a spatial autocorrelogram as the latter captures global pattern regularities. The distinct local grid patterns merged either abruptly or smoothly in the box interior. To quantify the amount of local pattern variation, we compared cross-correlations between quadrants in the 2.2-m box and the 1.5-m box. We could successfully capture the grid geometry in these smaller segments because we generated average quadrant autocorrelograms (from splitting each rate map into equal 2×2 sections) for each module. Cross-correlations were much higher in the 1.5-m box compared to the 2.2-m box, supporting the notion that the larger box induced local and more complex anchoring.

Finally, we performed the same analyses on the 2.2-m box data but with rate maps divided into 3×3 segments. There were clear differences in deformation patterns across these segments. Grid scores (rotational symmetry) were significantly higher in the central bin compared to the peripheral bins. Corner segments showed a particularly high degree of deformation, and in one corner—the corner where all animals were released into the box—ellipse direction showed a remarkably low degree of variation.

The need to anchor internal representations of space to external frames is paramount for allocentric function. We have demonstrated that grids align to the environment in a systematic manner. We have also suggested that the alignment of the grid pattern can be used to counteract mislocation within geometrically ambiguous environments. Rats tested in spatial working memory tasks in rectangular environments make systematic errors in segments of the box that have rotationally equivalent geometry, even in the presence of polarizing cues (Cheng 1986), which suggests geometric confusion is a common issue in spatial representation, as is supported by similar effects found in several other species (Cheng 2008).

We hypothesize that border cells provide mechanistic links between the grid map and the external environment. Despite abundant visual landmarks in the recording rooms, modules, with few exceptions, aligned according to the geometry of the environment. There may be a special salience given to environment borders, as opposed to more point-like visual cues, because environmental borders are generally more dependable and have an orientation. Biegler and Morris (1993) found that rats only used landmarks within an environment to gauge distances if the landmarks were stable within that environment. Several studies have shown similar connections to environmental geometry in other cell types (Save et al. 2000; Knierim et al. 1995, 1998; Sharp et al. 1995; Etienne et al. 2004) but have also highlighted the fact that the system's use of landmarks for spatial representation can be changed experimentally through learning (Jeffery et al. 1997). The close match between observed alignment and the alignment that would maximally decorrelate population codes across segments of the environment suggests that there could be a competitive interaction between path integration signals and sensory resets, as observed previously for place cells in the hippocampus (Gothard et al. 1996; Redish et al. 2000).

Oblique Effect in Grids?

Discrimination and detection of visual stimuli are dependent on the relationship of the stimulus to the axes of the environment, a well-known effect known as the 'oblique effect' (Mach 1860). Stimuli oriented along the cardinal axes yield better psychophysical performance compared to obliquely oriented stimuli. In the visual cortices, both single-neuron responses and population responses reflect this psychophysical anisotropy by increased representational density along the cardinal axes (Furmanski and Engel 2000; Wang et al. 2003; Xu et al. 2006). Several studies

suggest that the oblique effect originates in higher order cortices (Nasr and Tootell 2012; Liang et al. 2007; Shen et al. 2008), as the effect is stronger here compared to early sensory cortex (Shen et al. 2008; Müller et al. 2000), and the effect in early cortex is selectively abolished by temporal inactivation of higher order cortex (Shen et al. 2008). Grid cells are typically aligned close to parallel to the cardinal axes of the environment. Recently, it was shown that grid representations are not limited to navigational space in that a grid map of visual space was demonstrated in the entorhinal cortex of monkeys (Killian et al. 2012). Although highly speculative, it is interesting to ponder the possibilities for similar mechanisms at play in embedding internal representations into external reference frames in the visual domain as in the spatial domain. Although not very many examples were given by Killian et al. (2012), there seems to also be a trend for grids to align with a slight offset to cardinal axes (see their Fig. 1). Further, using optical imaging in area MT (which shows movement and orientation selectivity for stimuli) in the visual system, Xu et al. (2006) showed frequency plots of activation over the range of possible stimulus orientations. In these plots, there are quite distinct peaks with bimodal offsets from the cardinal axes (Fig. 7). Upon further inspection, these offsets are very close to 7.5° , which is the exact peak we observed in the alignment offset in grid cells. This finding points to a possible, albeit suppositional, link between visual

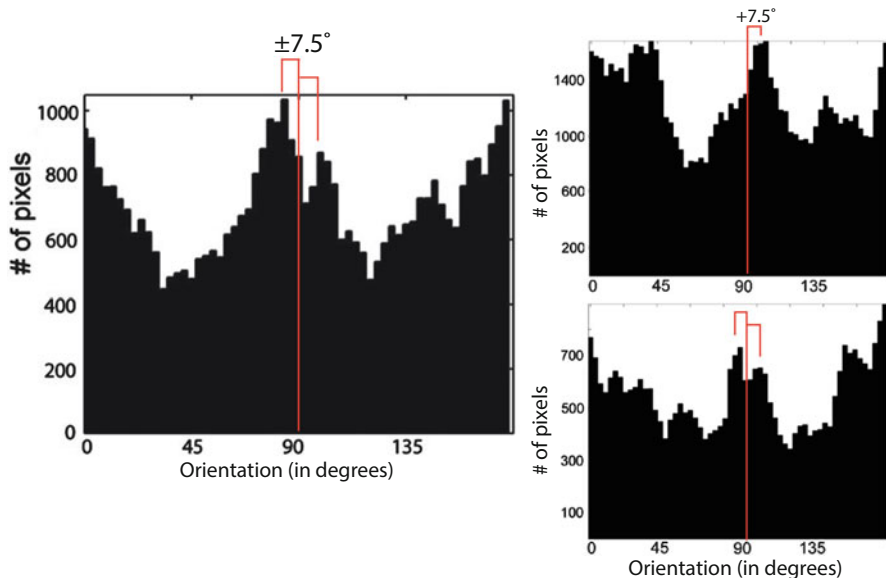


Fig. 7 The oblique effect in visual area MT in the owl monkey. Histograms show local activity measured by intrinsic optical imaging. Increased pixel count (y-axis) corresponds to higher activation. The different panels are from distinct subareas within MT. The *red lines* show 7.5° offsets calculated from the x-axis of the plots. Note the correspondence between peak offset and *red lines*. Reproduced with permission from Xu et al. (2006) (their Fig. 3 and Supplementary Fig. 6)

and spatial encoding in relation to real world axes, a link to be explored through future studies.

Conclusions

The entorhinal-hippocampal circuit offers a good model system for investigating basic functions carried out by neural networks within a behavioral context. Our understanding of grid cells has developed substantially at the single cell level but has for a while lagged behind the population insights gained from the hippocampus. By overcoming technical and analytic hurdles, we have now defined the first large-scale population characteristics of two central space-encoding cell populations. The grid map was shown to be modular, with considerable independence in the response of modules to geometrical features of the environment. Grid modules were found to use a general strategy to anchor grid orientation to the environment, pointing to this strategy as an optimal mechanism for population encoding of ambiguous segments of the external environment.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Barry C, Lever C, Hayman R, Hartley T, Burton S, O'Keefe J, Jeffery K, Burgess N (2006) The boundary vector cell model of place cell firing and spatial memory. *Rev Neurosci* 17:71–97
- Barry C, Hayman R, Burgess N, Jeffery KJ (2007) Experience-dependent rescaling of entorhinal grids. *Nat Neurosci* 10:682–684
- Bi GQ, Poo MM (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci* 18:10464–10472
- Biegler R, Morris RG (1993) Landmark stability is a prerequisite for spatial but not discrimination learning. *Nature* 361:631–633
- Brun VH, Otnass MK, Molden S, Steffenach HA, Witter MP, Moser MB, Moser EI (2002) Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. *Science* 296:2243–2246
- Brun VH, Solstad T, Kjelstrup KB, Fyhn M, Witter MP, Moser EI, Moser MB (2008) Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus* 18:1200–1212
- Burak Y, Fiete IR (2009) Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput Biol* 5:e1000291

- Burgalossi A, Herfst L, von Heimendahl M, Forste H, Haskic K, Schmidt M, Brecht M (2011) Microcircuits of functionally identified neurons in the rat medial entorhinal cortex. *Neuron* 70:773–786
- Burgess N, Barry C, O’Keefe J (2007) An oscillatory interference model of grid cell firing. *Hippocampus* 17:801–812
- Cheng K (1986) A purely geometric module in the rat’s spatial representation. *Cognition* 23:149–178
- Cheng K (2008) Whither geometry? Troubles of the geometric module. *Trends Cogn Sci* 12:355–361
- Couey JJ, Witoelar A, Zhang SJ, Zheng K, Ye J, Dunn B, Czajkowski R, Moser MB, Moser EI, Roudi Y, Witter MP (2013) Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat Neurosci* 16:318–324
- Derdikman D, Whitlock JR, Tsao A, Fyhn M, Hafting T, Moser MB, Moser EI (2009) Fragmentation of grid cell maps in a multicompartment environment. *Nat Neurosci* 12:1325–1332
- Dolorfo CL, Amaral DG (1998) Entorhinal cortex of the rat: topographic organization of the cells of origin of the perforant path projection to the dentate gyrus. *J Comp Neurol* 398:25–48
- Dombeck DA, Harvey CD, Tian L, Looger LL, Tank DW (2010) Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nat Neurosci* 13:1433–1440
- Dunn B, Mørreaunet M, Roudi Y (2015) Correlations and functional connections in a population of grid cells. *PLoS Comput Biol* 11:e1004052
- Etienne AS, Maurer R, Boukens V, Levy A, Rowe T (2004) Resetting the path integrator: a basic condition for route-based navigation. *J Exp Biol* 207:1491–1508
- Evensmoen HR, Lehn H, Xu J, Witter MP, Nadel L, Håberg AK (2013) The anterior hippocampus supports a coarse, global environmental representation and the posterior hippocampus supports fine-grained, local environmental representations. *J Cogn Neurosci* 25:1908–1925
- Fanselow MS, Dong H-W (2010) Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron* 65:7–19
- Fuhs MC, Touretzky DS (2006) A spin glass model of path integration in rat medial entorhinal cortex. *J Neurosci* 26:4266–4276
- Furmanski CS, Engel SA (2000) An oblique effect in human primary visual cortex. *Nat Neurosci* 3:535–536
- Fyhn M, Molden S, Witter MP, Moser EI, Moser M-B (2004) Spatial representation in the entorhinal cortex. *Science* 305:1258–1264
- Fyhn M, Hafting T, Treves A, Moser M-B, Moser EI (2007) Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* 446:190–194
- Garden DL, Dodson PD, O’Donnell C, White MD, Nolan MF (2008) Tuning of synaptic integration in the medial entorhinal cortex to the organization of grid cell firing fields. *Neuron* 60:875–889
- Giocomo LM, Hasselmo ME (2008) Time constants of h current in layer ii stellate cells differ along the dorsal to ventral axis of medial entorhinal cortex. *J Neurosci* 28:9414–9425
- Giocomo LM, Hasselmo ME (2009) Knock-out of HCN1 subunit flattens dorsal-ventral frequency gradient of medial entorhinal neurons in adult mice. *J Neurosci* 29:7625–7630
- Giocomo LM, Zilli EA, Fransén E, Hasselmo ME (2007) Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science* 315:1719–1722
- Giocomo LM, Hussaini SA, Zheng F, Kandel ER, Moser MB, Moser EI (2011) Grid cells use HCN1 channels for spatial scaling. *Cell* 147:1159–1170
- Giocomo LM, Stensola T, Bonnevie T, Van Cauter T, Moser M-B, Moser EI (2014) Topography of head direction cells in medial entorhinal cortex. *Curr Biol* 24:252–262
- Gothard KM, Skaggs WE, McNaughton BL (1996) Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues. *J Neurosci* 16:8027–8040
- Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI (2005) Microstructure of a spatial map in the entorhinal cortex. *Nature* 436:801–806

- Hafting T, Fyhn M, Bonnevie T, Moser M-B, Moser EI (2008) Hippocampus-independent phase precession in entorhinal grid cells. *Nature* 453:1248–1252
- Hargreaves EL, Rao G, Lee I, Knierim JJ (2005) Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science* 308:1792–1794
- Hartley T, Burgess N, Lever C, Cacucci F, O'Keefe J (2000) Modeling place fields in terms of the cortical inputs to the hippocampus. *Hippocampus* 10:369–379
- Harvey CD, Collman F, Dombeck DA, Tank DW (2009) Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* 461:941–946
- Henke PG (1990) Hippocampal pathway to the amygdala and stress ulcer development. *Brain Res Bull* 25:691–695
- Heys JG, Rangarajan KV, Dombeck DA (2014) The functional micro-organization of grid cells revealed by cellular-resolution imaging. *Neuron* 84:1079–1090
- Ikeda J, Mori K, Oka S, Watanabe Y (1989) A columnar arrangement of dendritic processes of entorhinal cortex neurons revealed by a monoclonal antibody. *Brain Res* 505:176–179
- Jeewajee A, Barry C, O'Keefe J, Burgess N (2008) Grid cells and theta as oscillatory interference: electrophysiological data from freely moving rats. *Hippocampus* 18:1175–1185
- Jeffery KJ, Donnett JG, Burgess N, O'Keefe JM (1997) Directional control of hippocampal place fields. *Exp Brain Res* 117:131–142
- Jung MW, Wiener SI, McNaughton BL (1994) Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *J Neurosci* 14:7347–7356
- Kaas JH (2012) Evolution of columns, modules, and domains in the neocortex of primates. *Proc Natl Acad Sci USA* 109(Suppl):10655–10660
- Killian NJ, Jutras MJ, Buffalo EA (2012) A map of visual space in the primate entorhinal cortex. *Nature* 491:761–764
- Kjelstrup KG, Tuvnes FA, Steffenach HA, Murison R, Moser EI, Moser MB (2002) Reduced fear expression after lesions of the ventral hippocampus. *Proc Natl Acad Sci USA* 99:10825–10830
- Kjelstrup KB, Solstad T, Brun VH, Hafting T, Leutgeb S, Witter MP, Moser EI, Moser MB (2008) Finite scale of spatial representation in the hippocampus. *Science* 321:140–143
- Knierim JJ, Kudrimoti HS, McNaughton BL (1995) Place cells, head direction cells, and the learning of landmark stability. *J Neurosci* 15:1648–1659
- Knierim JJ, Kudrimoti HS, McNaughton BL (1998) Interactions between idiothetic cues and external landmarks in the control of place cells and head direction cells. *J Neurophysiol* 80:425–446
- Krupic J, Burgess N, O'Keefe J (2012) Neural representations of location composed of spatially periodic bands. *Science* 337:853–857
- Langston RF, Ainge JA, Couey JJ, Canto CB, Bjerknes TL, Witter MP, Moser EI, Moser MB (2010) Development of the spatial representation system in the rat. *Science* 328:1576–1580
- Liang Z, Shen W, Shou T (2007) Enhancement of oblique effect in the cat's primary visual cortex via orientation preference shifting induced by excitatory feedback from higher-order cortical area 21a. *Neuroscience* 145:377–383
- Mach E (1860) Ueber das Sehen von Lagen und Winkeln durch die Bewegung des Auges. *Sitzungsberichte der Math Cl der Kais Akad der Wissenschaften* 42:215–224
- McNaughton BL, Battaglia FP, Jensen O, Moser EI, Moser MB (2006) Path integration and the neural basis of the 'cognitive map'. *Nat Rev Neurosci* 7:663–678
- Monaco JD, Abbott LF (2011) Modular realignment of entorhinal grid cell activity as a basis for hippocampal remapping. *J Neurosci* 31:9414–9425
- Montagnini A, Treves A (2003) The evolution of mammalian cortex, from lamination to arealization. *Brain Res Bull* 60:387–393
- Moser MB, Moser EI (1998) Functional differentiation in the hippocampus. *Hippocampus* 8:608–619
- Moser E, Moser M-B, Andersen P (1993) Spatial learning impairment parallels the magnitude of dorsal hippocampal lesions, but is hardly present following ventral lesions. *J Neurosci* 13:3916–3925

- Moser M-B, Moser EI, Forrest E, Andersen P, Morris RG (1995) Spatial learning with a minislab in the dorsal hippocampus. *Proc Natl Acad Sci USA* 92:9697–9701
- Moser EI, Roudi Y, Witter MP, Kentros C, Bonhoeffer T, Moser MB (2014) Grid cells and cortical representation. *Nat Rev Neurosci* 15:466–481
- Muller RU, Kubie JL (1987) The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J Neurosci* 7:1951–1968
- Müller T, Stetter M, Hübener M, Sengpiel F, Bonhoeffer T, Gödecke I, Chapman B, Löwel S, Obermayer K (2000) An analysis of orientation and ocular dominance patterns in the visual cortex of cats and ferrets. *Neural Comput* 12:2573–2595
- Nadel L (1968) Dorsal and ventral hippocampal lesions and behavior. *Physiol Behav* 3:891–900
- Nasr S, Tootell RB (2012) A cardinal orientation bias in scene-selective visual cortex. *J Neurosci* 32:14921–14926
- O’Keefe J (1976) Place units in the hippocampus of the freely moving rat. *Exp Neurol* 51:78–109
- O’Keefe J, Burgess N (1996) Geometric determinants of the place fields of hippocampal neurons. *Nature* 381:425–428
- O’Keefe J, Burgess N (2005) Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus* 15:853–866
- O’Keefe J, Conway FH (1978) Hippocampal place units in the freely moving rat: why they fire where they fire. *Exp Brain Res* 31:573–590
- O’Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res* 34:171–175
- O’Keefe J, Nadel L (1978) *The hippocampus as a cognitive map*. Oxford University Press, Oxford
- Poppenk J, Evensmoen HR, Moscovitch M, Nadel L (2013) Long-axis specialization of the human hippocampus. *Trends Cogn Sci* 17:230–240
- Rasmussen T, Penfield W (1947) The human sensorimotor cortex as studied by electrical stimulation. *Fed Proc* 6:184
- Redish AD, Rosenzweig ES, Bohanick JD, McNaughton BL, Barnes CA (2000) Dynamics of hippocampal ensemble activity realignment: time versus space. *J Neurosci* 20:9298–9309
- Redish AD, Battaglia FP, Chawla MK, Ekstrom AD, Gerrard JL, Lipa P, Rosenzweig ES, Worley PF, Guzowski JF, McNaughton BL, Barnes CA (2001) Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. *J Neurosci* 21:RC134
- Samsonovich A, McNaughton BL (1997) Path integration and cognitive mapping in a continuous attractor neural network model. *J Neurosci* 17:5900–5920
- Save E, Nerad L, Poucet B (2000) Contribution of multiple sensory information to place field stability in hippocampal place cells. *Hippocampus* 10:64–76
- Savelli F, Yoganarasimha D, Knierim JJ (2008) Influence of boundary removal on the spatial representations of the medial entorhinal cortex. *Hippocampus* 18:1270–1282
- Sharp E, Blair HT, Etkin D, Tzanetos B (1995) Influences of vestibular and visual motion information on the spatial firing patterns of hippocampal place cells. *J Neurosci* 15:173–189
- Shen W, Liang Z, Shou T (2008) Weakened feedback abolishes neural oblique effect evoked by pseudo-natural visual stimuli in area 17 of the cat. *Neurosci Lett* 437:65–70
- Solodkin A, Van Hoesen GW (1996) Entorhinal cortex modules of the human brain. *J Comp Neurol* 365:610–617
- Solstad T, Boccara CN, Kropff E, Moser M-B, Moser EI (2008) Representation of geometric borders in the entorhinal cortex. *Science* 322:1865–1868
- Stensola H, Stensola T, Solstad T, Frøland K, Moser MB, Moser EI (2012) The entorhinal grid map is discretized. *Nature* 492:72–78
- Stensola T, Stensola H, Moser M-B, Moser EI (2015) Shearing-induced asymmetry in entorhinal grid cells. *Nature* 518:207–212
- Tolman EC (1948) Cognitive maps in rats and men. *Psychol Rev* 55:189–208
- Van Cauter T, Poucet B, Save E (2008) Unstable CA1 place cell representation in rats with entorhinal cortex lesions. *Eur J Neurosci* 27:1933–1946
- Wang G, Ding S, Yunokuchi K (2003) Difference in the representation of cardinal and oblique contours in cat visual cortex. *Neurosci Lett* 338:77–81

- Welinder PE, Burak Y, Fiete IR (2008) Grid cells: the position code, neural network models of activity, and the problem of learning. *Hippocampus* 18:1283–1300
- Wills TJ, Cacucci F, Burgess N, O’Keefe J (2010) Development of the hippocampal cognitive map in preweanling rats. *Science* 328:1573–1576
- Wilson MA, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. *Science* 261:1055–1058
- Witter MP, Groenewegen HJ, Lopes da Silva FH, Lohman AH (1989) Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region. *Prog Neurobiol* 33:161–253
- Xu X, Collins CE, Khaytin I, Kaas JH, Casagrande VA (2006) Unequal representation of cardinal vs. oblique orientations in the middle temporal visual area. *Proc Natl Acad Sci USA* 103:17490–17495
- Zhang SJ, Ye J, Miao C, Tsao A, Cerniauskas I, Ledergerber D, Moser MB, Moser EI (2013) Optogenetic dissection of entorhinal-hippocampal functional connectivity. *Science* 340:1232627

The Striatum and Decision-Making Based on Value

Ann M. Graybiel

Abstract Our behaviors range from mindful, deliberative streams of action to sequences of action that are so nearly automatic that we can perform them almost without thinking. Transitions between these modes of behavior occur as we learn behavioral routines. We have studied these transitions and the neural activity that occurs in corticostriatal loops as they take place. We find that neural activity in these loops is strongly modified during habit learning and that specific corticostriatal circuits can powerfully control value-based decision-making and habits.

As we move about and act in our environment, the brain constantly updates not only our physical position and the moment-to-moment stimuli around us, but also updates the value of the actions that we perform. How these values are attached to our behaviors is still incompletely understood.

In our laboratory, we have approached this issue by teaching animals to perform simple habits, capitalizing on much evidence that, at first, behaviors that are candidate habits are sensitive to reinforcement, but later they become nearly independent of whether or not the performance of the behavior is reinforced.

We have found that as this behavioral transition occurs, the spike activity and local field potential activity recorded in the prefrontal cortex and striatum are also transformed (Jog et al. 1999; Barnes et al. 2005; Thorn et al. 2010; Smith and Graybiel 2013). In typical experiments, we have taught rodents to run in simple T-mazes, with cues indicating to them whether to turn left or right to receive a food reward. The neural activity in regions known to be necessary for habit formation gradually shifts: early on, the population activity in the sensorimotor part of the striatum is high during the full time of the maze runs, but later during the learning process, the population activity becomes concentrated at the action points of the

A.M. Graybiel (✉)

McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

e-mail: graybiel@mit.edu

© The Author(s) 2016

G. Buzsáki, Y. Christen (eds.), *Micro-, Meso- and Macro-Dynamics of the Brain*, Research and Perspectives in Neurosciences, DOI 10.1007/978-3-319-28802-4_6

runs, especially the beginning and end of the runs. As the behavior of the animals becomes fully habitual through extensive training (called ‘over-training’) on the task, this beginning-and-end bracketing pattern becomes nearly fixed within the sensorimotor striatum. A quite similar bracketing pattern later develops in the prefrontal cortex, but it remains sensitive to reinforcement; if rewards are made unpalatable, then the animals cease the habitual runs and the cortical bracketing activity pattern becomes degraded.

We then found that we could block already formed habits and even toggle the habit off and on by optogenetically suppressing this prefrontal cortical activity (Smith et al. 2012). Comparable optogenetic inhibition of the same small prefrontal cortical zone could block the formation of habits altogether when the optogenetic inhibition was applied during the over-training period (Smith and Graybiel 2013).

These experiments raise the possibility that neural circuits involving the medial prefrontal cortex can evaluate whether actions are beneficial and should be allowed to be performed. The fact that this apparent control is effective even for behaviors that seem to be nearly fully automatic suggests that there is on-line, value-related control of behavior.

This potential was vividly seen in other experiments in which we blocked compulsive grooming behavior in a mouse model of obsessive-compulsive disorder by manipulating an orbitofrontal corticostriatal circuit (Burguiere et al. 2013). In these experiments, we could block a conditioned compulsion by intervening either at the level of the cortex or at the level of the medial striatum. Therefore, the control was exerted by a corticostriatal circuit.

In a new set of experiments, we have asked whether we can identify critical corticostriatal circuits that operate in these deliberative or repetitive decisions. We focused on a circuit that is thought to lead from localized zones in the prefrontal cortex to striosomes. These are dispersed zones within the striatum that can access the dopamine-containing neurons of the midbrain (Crittenden and Graybiel 2011; Fujiyama et al. 2011; Watabe-Uchida et al. 2012). We mimicked a situation often faced in everyday life, in which we can acquire something, but only at a cost. In this situation, costs as well as benefits have to be weighed. We used decision-making tasks in which animals were required to choose an action sequence in response to cues indicating that mixtures of rewarding and annoying reinforcers could either be accepted or be rejected. This design meant that the animals could reject an offer, but then they would miss out on the reward coupled to the cost.

This kind of decision-making, given the name ‘approach-avoidance decision-making,’ has been studied extensively in human subjects, particularly in relation to distinguishing between anxiety and depression in affected individuals who face conflicting motivations to approach and to avoid. We thus were attempting to target forms of decision-making that, in humans, involve value-based estimates of the future.

In initial studies, Dr. Ken-ichi Amemori and I focused on the pregenual anterior cingulate cortex in macaque monkeys (Amemori and Graybiel 2012), which earlier work had shown to project preferentially to striosomes in the head of the caudate nucleus (Eblen and Graybiel 1995). There, many neurons increased their activity

during the decision period, either when the monkey would subsequently choose an approach response (accepting the good and bad symbolized by cues on a computer screen) or when the monkey would subsequently reject the offer. In one localized pregenual region, the avoidance-related neurons outnumbered the approach-related neurons. At other sites, similar numbers of these two classes were recorded. Microstimulation applied during the decision period had little or no effect on the decisions at most sites, but in the regions matching the sites with predominance of avoidance-related neurons, the microstimulation induced significant increases in avoidance. We found that treatment with the anxiolytic diazepam could block the microstimulation effects. Notably, we found no effects of the microstimulation in a control ‘approach-approach’ task in which both offered options were good.

In subsequent, still-ongoing experiments, Ken-ichi Amemori, Satoko Amemori and I are determining whether, as initial results suggest, the ‘hot-spot’ for pessimistic decision-making preferentially projects to striosomes (Amemori et al. in preparation). If so, these experimental findings would squarely place the corticostriatal system interacting with striosomes as part of the circuitry underpinning decision-making in which conflicting motivations must be handled.

With the technical opportunities presented by work in rodents, we returned to T-maze experiments, but this time introduced costs and benefits at each end-arm of the mazes. In work spearheaded by Alexander Friedman, Daigo Homma, and Leif Gibb, with Ken-ichi Amemori and others, we found striking evidence for a selective functional engagement of a striosome-targeting prefrontal circuit (Friedman et al. 2015). The evidence rests on the use of multiple decision-making tasks, presenting cost-benefit, benefit-benefit, reverse cost-benefit and cost-cost decision-making challenges to the animals. We then used optogenetics to interrupt the cortico-striosomal circuit. Across all of these tasks, it was only in the cost-benefit task that the putative striosome-targeting prefrontal pathway was engaged. By contrast, comparable optogenetic experiments inhibiting a matrix-targeting prefronto-striatal circuit produced effects on decision-making in all of the tasks.

Evidence from our own and other laboratories suggests that striosomes may have privileged access to the dopamine-containing neurons of the substantia nigra pars compacta, either directly or by way of a multi-synaptic pathway via the lateral habenula (Rajakumar et al. 1993; Graybiel 2008; Stephenson-Jones et al. 2013). The details of these pathways remain unknown. It is known, however, that the lateral habenula neurons increase their firing rates to negative reinforcers or their predictors; the dopamine-containing nigral neurons fire in relation to positive or, in some populations, to negative reinforcers and predictors (Hong and Hikosaka 2013). This potential dual downstream circuitry, combined with the experimental evidence summarized here, suggests that striosomes could be nodal sites in mood- and emotion-related corticostriatal networks influencing downstream modulators of motivational states.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any

noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Amemori K, Graybiel AM (2012) Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. *Nat Neurosci* 15:776–785
- Barnes T, Kubota Y, Hu D, Jin DZ, Graybiel AM (2005) Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* 437:1158–1161
- Burguiere E, Monteiro P, Feng G, Graybiel AM (2013) Optogenetic stimulation of lateral orbitofronto-striatal pathway suppresses compulsive behaviors. *Science* 340:1243–1246
- Crittenden JR, Graybiel AM (2011) Basal ganglia disorders associated with imbalances in the striatal striosome and matrix compartments. *Front Neuroanat* 5:59
- Eblen F, Graybiel AM (1995) Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. *J Neurosci* 15:5999–6013
- Friedman A, Homma D, Gibb LG, Amemori K, Rubin SJ, Hood AS, Riad MH, Graybiel AM (2015) A corticostriatal path targeting striosomes controls decision-making under conflict. *Cell* 161:1320–1333
- Fujiyama F, Sohn J, Nakano T, Furuta T, Nakamura KC, Matsuda W, Kaneko T (2011) Exclusive and common targets of neostriatofugal projections of rat striosome neurons: a single neuron-tracing study using a viral vector. *Eur J Neurosci* 33:668–677
- Graybiel AM (2008) Habits, rituals and the evaluative brain. *Annu Rev Neurosci* 31:359–387
- Hong S, Hikosaka O (2013) Diverse sources of reward value signals in the basal ganglia nuclei transmitted to the lateral habenula in the monkey. *Front Hum Neurosci* 7:778
- Jog M, Kubota Y, Connolly CI, Hillegaart V, Graybiel AM (1999) Building neural representations of habits. *Science* 286:1745–1749
- Rajakumar N, Elisevich K, Flumerfelt BA (1993) Compartmental origin of the striato-entopeduncular projection in the rat. *J Comp Neurol* 331:286–296
- Smith KS, Graybiel AM (2013) A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron* 79:361–374
- Smith KS, Virkud A, Deisseroth K, Graybiel AM (2012) Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. *Proc Natl Acad Sci USA* 109:18932–18937
- Stephenson-Jones M, Kardamakis AA, Robertson B, Grillner S (2013) Independent circuits in the basal ganglia for the evaluation and selection of actions. *Proc Natl Acad Sci USA* 110:E3670–E3679
- Thorn CA, Atallah H, Howe M, Graybiel A (2010) Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron* 66:781–795
- Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A, Uchida N (2012) Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74:858–873

Decoding the Dynamics of Conscious Perception: The Temporal Generalization Method

Stanislas Dehaene and Jean-Rémi King

Abstract Parsing a cognitive task into a sequence of successive operations is a central problem in cognitive neuroscience. A major advance is now possible thanks to the application of pattern classifiers to time-resolved recordings of brain activity [electro-encephalography (EEG), magneto-encephalography (MEG), or intracranial recordings]. The method determines precisely when a specific mental content becomes explicitly represented in brain activity. Most importantly, the ability of these pattern classifiers to generalize across time and experimental conditions sheds light on the temporal organization of information-processing stages. To illustrate these ideas, we show how the decoding of MEG and EEG recordings can be used to track the fate of conscious and unconscious brain processes during simple masking and auditory novelty tasks. The experimental results yield converging results, suggesting that conscious perception is associated with the late formation of a distributed and stable neural assembly that encodes the content of subjective perception.

Parsing a cognitive task into a sequence of successive operations has been recognized as a central problem ever since the inception of scientific psychology. The Dutch ophthalmologist Franciscus Donders first used mental chronometry to demonstrate that mental operations are slow and can be decomposed into a series of successive stages (Donders 1969). Since then, psychologists have proposed a variety of elegant but indirect methods by which such decomposition could be achieved using behavioral measurements of response times (Pashler 1994; Posner 1978; Sigman and Dehaene 2005; Sternberg 1969, 2001).

The American psychologist and cognitive neuroscientist Michael Posner was among the first to realize that the advent of brain imaging methods provided direct evidence of this classical task-decomposition problem, and he successfully

S. Dehaene (✉)
Collège de France, Paris, France

INSERM-CEA Cognitive Neuroimaging Unit, NeuroSpin Center, Saclay, France
e-mail: stanislas.dehaene@gmail.com

J.-R. King
INSERM-CEA Cognitive Neuroimaging Unit, NeuroSpin Center, Saclay, France

analyzed several tasks such as reading or attention orienting into their component operations (Petersen et al. 1988; Posner and Raichle 1994). Time-resolved methods that capture brain activity at the scale of milliseconds, such as electro- and magneto-encephalography (EEG and MEG) or intracranial recordings, seem particularly well suited to this task-decomposition problem, because they can reveal how the brain activity unfolds over time in different brain areas, each potentially associated with a specific neural code. Yet the amount and the complexity of electrophysiological recordings can rapidly become overwhelming. In particular, it remains difficult to accurately reconstruct the spatial sources of EEG and MEG signals. As a result, the series of operations underlying basic cognitive tasks remain ill-defined in most cases.

Machine learning techniques, combined with high-temporal-resolution brain imaging methods, now provide a new tool with which to address this question. In this chapter, we briefly review a technique that we call the “temporal generalization method” (King and Dehaene 2014), which clarifies how multiple processing stages and their corresponding neural codes unfold over time. We illustrate this method with several examples, and we use them to draw some conclusions about the dynamics of conscious processing.

The Temporal Generalization Method

Contemporary brain imaging techniques such as EEG and MEG typically allow us to simultaneously record a large number of electrophysiological signals from the healthy human brain (e.g., 256 sensors in EEG and 306 sensors in MEG). Similarly, using intracranial electrodes in monkeys or in human patients suffering from epilepsy, hundreds of electrophysiological signals can be acquired at rates of 1 kHz or above. Identifying, from such multidimensional signals, the neuronal representations and computations explicitly recruited at each processing stage can be particularly difficult. For example, reconstructing the neural source of EEG and MEG signals—i.e., determining precisely where in the brain the signals originate—remains a major hurdle. Signals from multiple areas are often superimposed in the recordings from a given sensor and, conversely, the signal from a given brain area simultaneously projects onto multiple sensors.

Machine learning techniques can help overcome these difficulties (Fig. 1). The idea is to provide a time slice of electrophysiological signals to a machine-learning algorithm that learns to extract, from this raw signal, information about a specific aspect of the stimulus. For instance, one can ask the algorithm to look for information about whether the visual stimulus was a vertical or a horizontal bar, whether a sound was rare or frequent, whether the subject responded with the right or the left hand, etc. If we train one such classifier for each time point t (or for a time window centered on time t), we obtain a series of classifiers whose performance traces a curve that tells us how accurately the corresponding parameter can be decoded at each moment in time. Typically, this curve remains at chance level before the onset of the stimulus, then quickly rises, and finally decays (Fig. 1).

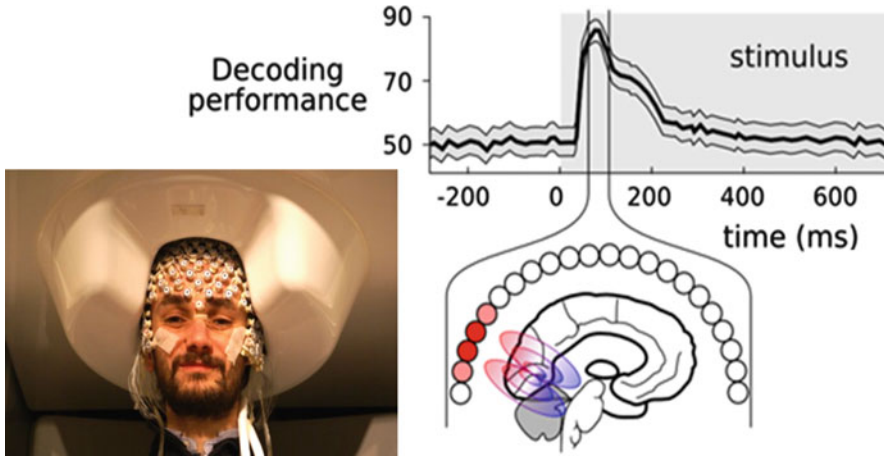


Fig. 1 Principle of temporal decoding (from King 2014). On each trial we simultaneously recorded a large number of brain signals (e.g., 256 EEG and 306 MEG signals). Using the data from a single time point t , or from a time window centered on time t , we could train a Support Vector Machine (SVM) to decode one aspect of the stimulus (for instance, the orientation of a grid on the subject's retina). The time course of decoding performance reveals the dynamics with which the information is represented in the brain. How a decoder trained at time t generalizes to data from another time point, t' , reveals whether the neural code changes over time

The decoding curves tracking distinct features of the current trial typically rise and fall at different times, thus providing precious indications about when, and in which order, the respective representations begin to be explicitly coded in brain activity. For example, Fig. 2 illustrates how we decoded the time course of perceptual, motor, intentional and meta-cognitive error-detection processes from the very same MEG/EEG signal (Charles et al. 2014; for another application to the stages of invariant visual recognition, see Isik et al. 2014).

In addition to tackling the *when* question, machine learning may also tell us for *how long* a given neural code is activated and whether it recurs over time. To this aim, we asked how a pattern classifier trained at time t generalizes to data from another time point t' . This approach results in a temporal generalization matrix that contains a vast amount of detail about the dynamics of neural codes (King and Dehaene 2014). If the same neural code is active at times t and t' , then a classifier trained at time t should generalize to the other time, t' . If, however, the information is passed on to a series of successive stages, each with its own coding scheme, then such generalization across time should fail, and classifiers trained at different time points will be distinct from each other. More generally, the shape of the temporal generalization matrix, which encodes the success in training at time t and testing at time t' for all combinations of t and t' , can provide a considerable amount of information about the time course of coding stages. For instance, it can reveal whether and when a given neural code recurs, how long it lasts, and whether its scalp topography reverses or oscillates. When comparing two experimental conditions C and C' , it can also reveal whether and when the series of unfolding stages

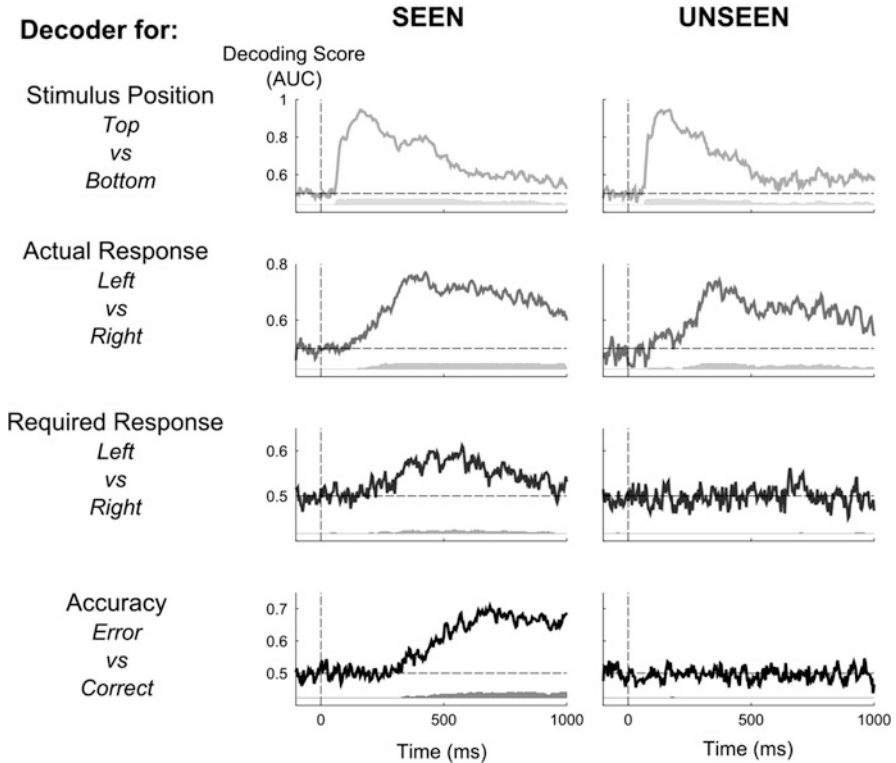


Fig. 2 Example of temporal decoding (from Charles et al. 2014). Distinct decoders were trained to extract four different properties of an unfolding trial from the same MEG and EEG signals: the position of a visual target on screen, the motor response made by the subject, the response that he should have made, and whether the response made was correct or erroneous. Note how those four distinct properties successively emerge in brain signals, from *left* to *right*. The target was masked, such that subjects occasionally reported it as “unseen” (*right column*). In this case, stimulus position and motor response could be decoded, but the brain seemed to fail to record either the required response or the accuracy of the motor response

was delayed, interrupted or reorganized (for detailed discussion, see King and Dehaene 2014).

Advantages of Multivariate Decoding Methods

Temporal decoding and temporal generalization are powerful multivariate methods that present several advantages over classical univariate methods for the characterization of brain activity:

- Within each subject, machine learning methods search for an optimal combination of brain signals that reflects a certain psychological variable. By combining

multiple sensors, the noise level can be drastically reduced, thus optimizing the detection of a significant effect. This technique is particularly useful when working with brain-lesioned patients in whom the topography of brain signals may be distorted; the software essentially replaces the experimenter in searching for significant brain signals (King et al. 2013a).

- “Double-dipping,” i.e., using the same data for inference and for confirmatory purposes, a problem that often plagues brain-imaging research (Kriegeskorte et al. 2009), can be largely circumvented in computer-based inference by leaving a subset of the data out of the training database and using it specifically to independently test the classifier.
- Hundreds of brain sensors are summarized into a single time curve that “projects” the data back onto a psychological space of interest. By identifying a near-optimal spatial filter, this aspect of the method simultaneously bypasses the complex problems of source reconstruction and of statistical correction for multiple comparisons across hundreds of sensors and provides cognitive scientists with immediately interpretable signals.
- Finally, because distinct classifiers are trained for different subjects, and only the projections back to psychological space are averaged across subjects, the method naturally takes into account inter-individual variability in brain topography. In this respect, the method makes fewer assumptions than classical univariate methods that implicitly rest on the dubious assumption that different subjects share a similar topography over EEG or MEG sensors. In the decoding approach, we do not average sensor-level data but only their projection onto a psychological dimension that is likely to be shared across subjects.

A drawback of the decoding method is that we cannot be sure that the features that we decode from brain signals are actually being used by the brain itself for its internal computations. For all we know, we could be decoding the brain’s equivalent of the steam cloud arising from a locomotive—a side effect rather than a causally relevant signal. To mitigate this problem, we restrict ourselves to the use of linear classifiers such as a linear Support Vector Machine (SVM). In this way, we can at least increase our confidence in the fact that the decoding algorithm focuses on *explicit* neural codes. A neural code for a feature f may be considered as “explicit” when f can be reconstructed from the neural signals using a simple linear transformation. For instance, the presence of faces versus other visual categories is explicitly represented in inferotemporal cortex because many neurons fire selectively to faces, and thus a simple averaging operation suffices to discriminate faces from non-faces (Tsao et al. 2006). This definition of “explicit representation” ensures that the brain has performed a sufficient amount of preprocessing to attain a level of representation that can be easily extracted and manipulated at the next stage of neural processing, either by single neurons or by neuronal assemblies. If we used sophisticated non-linear classifiers such as “deep” convolutional neural networks (LeCun et al. 2015), we could, at least in principle, decode any visual information from the primary visual area V1, but this would be uninformative about when, how and even whether the brain itself explicitly represents this

information. By using linear classifiers, we ensure that we only decode explicit neural signals. It should be kept in mind, however, that the identification of explicit representations with linearly separable ones is a working hypothesis that remains under-investigated. More generally, it is particularly difficult to determine whether and how brain responses play a causal role in behavior and subjective perception (see, e.g., Rangarajan et al. 2014). Beyond decoding analyses, this ambiguity is in fact intrinsic to any non-causal brain-behavior correlation method.

A Test Using Auditory Novelty Signals

Figure 3 illustrates an application of the temporal generalization method to auditory novelty detection in the local/global paradigm (Bekinschtein et al. 2009; King et al. 2013a). This paradigm aims to separate two types of brain signals evoked by the violation of two types of auditory expectations: (1) automatic detection of unexpected sounds, and (2) conscious detection of unexpected sound sequences. As we shall see, the temporal generalization analysis separates these two intermingled signals, facilitates their detection, and shows that their temporal dynamics differ radically.

We recorded MEG and EEG signals while human subjects heard sequences of five repeated sounds (Wacongne et al. 2011). Sometimes the auditory sequence ended with a different sound. This unexpected local violation generated a local mismatch response, arising primarily from auditory cortex (Bekinschtein et al. 2009). Furthermore, in each block, the entire sequence was repeated several times and, occasionally, was violated by presenting a rare instance of a distinct sequence. The difference between rare and frequent sequences generated a global novelty response, arising from distributed brain areas including associative areas of parietal and prefrontal cortex, and associated with a P3b component of the event-related potential (Bekinschtein et al. 2009).

Temporal decoding allowed us to track the corresponding novelty signals in the brain. First, classifiers could be trained to discriminate whether the fifth sound was repeated or deviant (local mismatch). Above-chance decoding scores could be observed during a time window $\sim 100\text{--}400$ ms after the deviant sound. Crucially, the temporal generalization matrix revealed that this long period did not correspond to a single neural code (Fig. 3b). A diagonal generalization pattern indeed suggested that error signals changed over time as they propagated through a hierarchy of distinct brain areas. There were even periods of below-chance generalization (marked in blue in Fig. 3b), indicating that the spatial pattern of brain activity observed at time t tended to reverse at time t' , possibly due to top-down inputs to the same brain area that have been postulated to play a role in cancelling out the bottom-up error signals (Friston 2005).

Second, the global effect was marked by a completely distinct pattern of temporal generalization. From about 150 ms on, classifiers could discriminate whether the global sequence was frequent or rare. The results demonstrated a

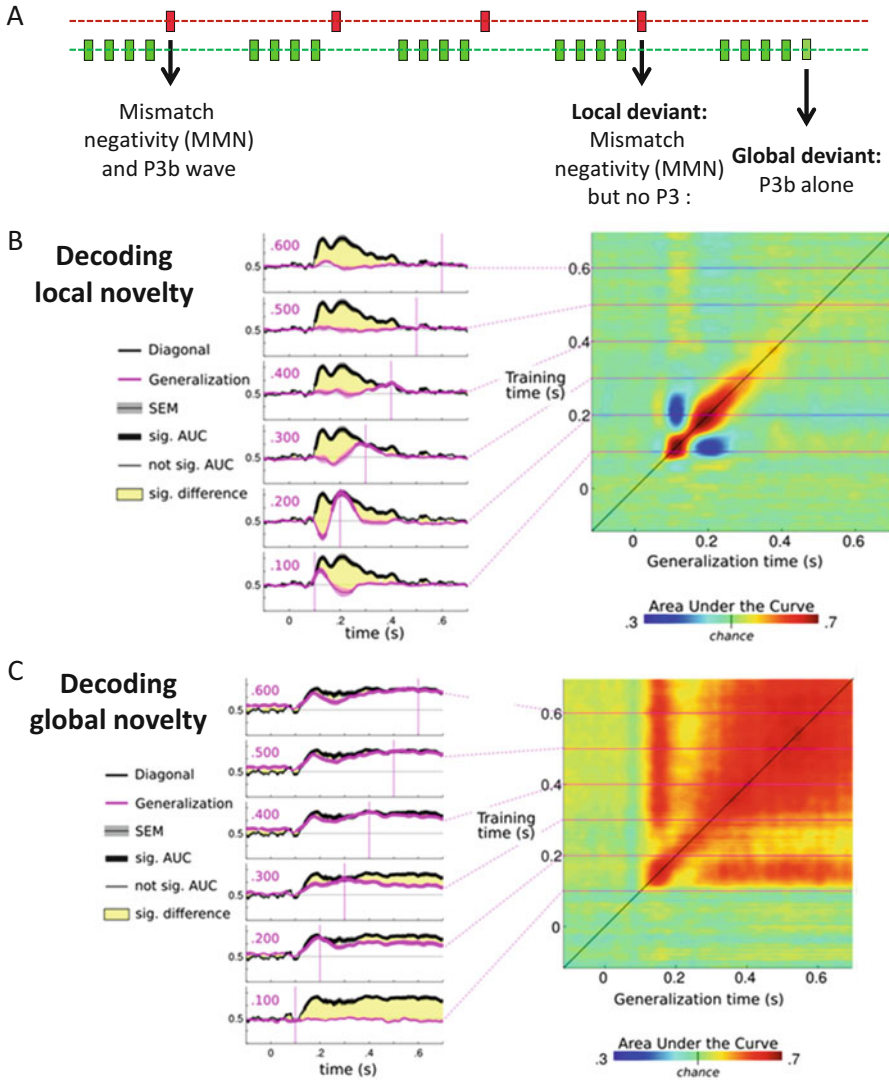


Fig. 3 Temporal decoding applied to an auditory violation paradigm, the local/global paradigm (from King et al. 2013a). (a) Experimental design: sequences of five sounds sometimes end with a different sound, generating a local mismatch response. Furthermore, the entire sequence is repeated and occasionally violated, generating a global novelty response (associated with a P3b component of the event-related potential). (b, c) Results using temporal decoding. A decoder for the local effect (b) is trained to discriminate whether the fifth sound is repeated or different. This is reflected in a diagonal pattern, suggesting the propagation of error signals through a hierarchy of distinct brain areas. Below-chance generalization (in blue) indicates that the spatial pattern observed at time t tends to reverse at time t' . A decoder for the global effect (c) is trained to discriminate whether the global sequence is frequent or rare. This is reflected primarily in a square pattern, indicating a stable neural pattern that extends to the next trial. In all graphs, $t = 0$ marks the onset of the fifth sound

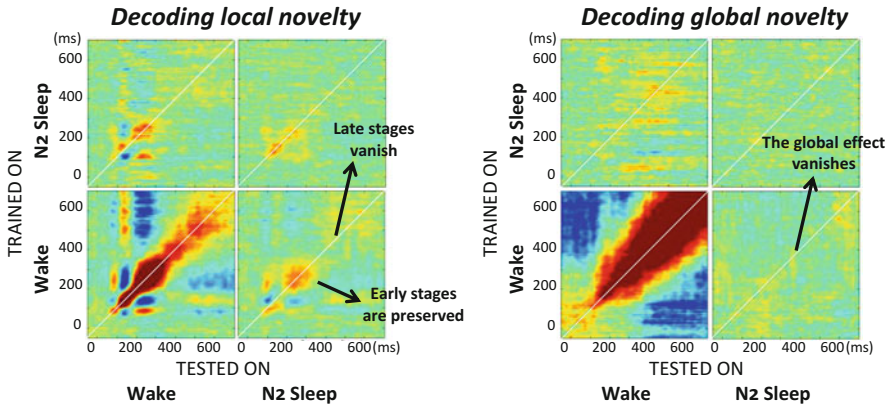


Fig. 4 Generalization of decoding across two experimental conditions, wakefulness and sleep, can reveal which processing stages are preserved or deleted (from Strauss et al. 2015). Subjects were tested with the same local/global paradigm as in Fig. 2 while they fell asleep in the MEG scanner. The local effect was partially preserved during sleep (*left*): between about 100 and 300 ms, a decoder could be trained during wake and generalize to sleep, or vice versa. Note that all late components and, interestingly, off-diagonal below-chance components vanished during sleep. As concerns the global effect (*right*), it completely vanished during sleep

square pattern of temporal generalization (Fig. 3c), indicating that the violation of global sequence expectations evoked a single and largely stable pattern of neural activity (with only a small enhancement on the diagonal, indicating a slow change in neural coding).

Further research showed that the late global response is a plausible marker of conscious processing (Dehaene and Changeux 2011): if processing reaches this level of complexity, whereby the present sequence is represented and compared to those heard several seconds earlier, then the person is consciously representing the deviant sequence and can later report it (Bekinschtein et al. 2009). Inattention abolishes the late global response but not the early local response. So does sleep: as soon as a person falls asleep and ceases to respond to the global deviants, the global response vanishes whereas the local response remains partially preserved, at least in its initial components (Fig. 4; see Strauss et al. 2015).

The disappearance of late and top-down processing stages seems to be a general characteristic of the loss of consciousness (for review, see Dehaene and Changeux 2011). In the local/global paradigm, when patients fall into a vegetative state or in a coma, the global effect vanishes whereas the local effect remains preserved. The global effect may therefore be used as a “signature” of conscious processing, useful to detect that consciousness is in fact preserved in a subset of patients in apparent vegetative state. In such patients, the temporal decoding method can optimize the detection of a global effect, even in the presence of delays or topographical distortions due to brain and skull lesions (King et al. 2013a). Unfortunately, the global effect is not a very sensitive signature of consciousness, because it may remain undetectable in some patients who are demonstrably conscious yet unable to

attend or whose EEG signals are contaminated by noise. When the global effect is present, however, it is likely that the patient is conscious or will quickly recover consciousness (Faugeras et al. 2011, 2012). Therefore, the decoding of the global effect adds to the panoply of recent EEG-based mathematical measures that, collectively, contribute to the accurate classification of disorders of consciousness in behaviorally unresponsive patients (King et al. 2013b; Sitt et al. 2014).

Late Metastable Activity as a Signature of Consciousness

Why does the global response to auditory novelty track conscious processing? We have hypothesized that conscious perception corresponds to the entry of information into a global neuronal workspace (GNW), based on distributed associative areas of the parietal, temporal and prefrontal cortices, that stabilizes information over time and broadcasts it to additional processing stages (Dehaene and Naccache 2001; Dehaene et al. 2003, 2006). Even if the incoming sensory information is very brief, the GNW transforms and stabilizes its representation for a period of a few hundreds of milliseconds, as long as is necessary to achieve the organism's current goals. Such a representation has been called "metastable" (Dehaene et al. 2003) by analogy with the physics of low-energy attractor states, where metastability is defined as "the phenomenon when a system spends an extended time in a configuration other than the system's state of least energy" (Wikipedia). Similarly, conscious representations are thought to rely on brain signals that persist for a long duration, yet without being fully stable because they can be suddenly replaced as soon as a new mental object becomes the focus of conscious thought.

The brain activity evoked by global auditory violations in the local/global paradigm fits with this hypothesis. First, this signal is only present in conscious subjects who can explicitly report the presence of deviant sequences. Furthermore, this signal is late, distributed in many high-level association areas including prefrontal cortex, and stable for an extended period of time (Bekinschtein et al. 2009). The latter point is particularly evident in temporal generalization matrices, which show that the global effect, although triggered by a transient auditory signal (a single 150-ms tone), is reflected in a late and approximately square (Fig. 3) or thick-diagonal (Fig. 4) pattern of decoding. Such a pattern indicates that the evoked neural pattern is stable over a long time period. Our results indicate that the neural activation pattern can be either quasi-stable for hundreds of milliseconds (as occurs in Fig. 3, where subjects simply had the instruction to attend to the stimuli), or slowly changing with considerable temporal overlap among successive neural codes (as occurs in Fig. 4, where subjects were instructed to perform a motor response to global deviants, thus enforcing a series of additional decision, response and monitoring stages).

Many additional paradigms have revealed that conscious access is associated with an amplification of incoming information, its transformation into a metastable representation, and its efficient propagation to subsequent processing stages (Del

Cul et al. 2007; Kouider et al. 2013; Salti et al. 2015; Schurger et al. 2015; Sergent et al. 2005). For example, Fig. 5 shows the results of temporal decoding applied to a classical masking paradigm, in which a digit is made invisible by following it at a short latency with a “mask” made up of letters surrounding the digit’s position (Charles et al. 2014; Del Cul et al. 2007). At short delays, subjects report the absence of a digit even when it is physically present on screen. Nevertheless, a pattern classifier can be trained to discriminate digit-present and digit-absent trials (thus decoding, from the subject’s brain, a piece of information that the subject himself ignores). The classifier for subliminal digits presents a sharp diagonal pattern (Fig. 5), indicating that the digit traverses a series of transient coding stages without ever stabilizing into a long-lasting activation. When the digit is seen, however, a square pattern of temporal generalization can be observed, suggesting a metastable representation of the digit’s presence. A similar difference in metastability can be observed when sorting physically identical threshold trials (SOA = 50 ms) into those that were subjectively reported as seen or unseen (Fig. 5).

Metastability can also be assessed by other means, for instance, by measuring whether the neural activation “vector” evoked by a given stimulus points in a consistent direction for a long-enough duration (Schurger et al. 2015). Here again, a few hundreds of milliseconds after the onset of a picture, stability was

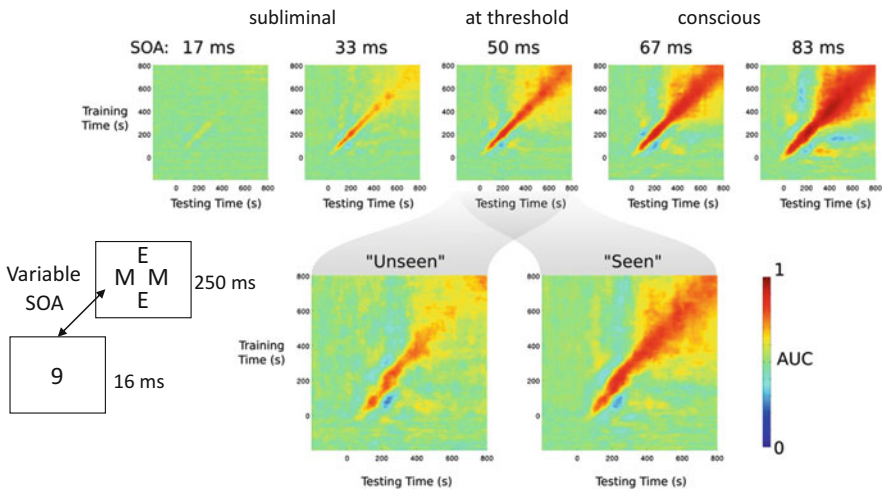


Fig. 5 Decoding reveals the signatures of subliminal and conscious processing in a masking paradigm (data from Charles et al. 2013, 2014). When the stimulus-onset-asynchrony (SOA) between a digit and a letter mask remains below 50 ms, the digit generally remains subjectively invisible. A decoder trained to discriminate digit-present and digit-absent trials decodes only a sharp diagonal pattern, indicating that the digit quickly traverses a series of successive coding stages. When the digit is seen, however, a square pattern of temporal generalization emerges, indicating that a temporally stable representation is achieved. A similar, though more modest difference, can be observed when sorting physically identical threshold trials (SOA = 50 ms) into those that were subjectively reported as seen or unseen

higher when the picture was consciously perceived than when it was unseen. Thus, late metastability consistently appears to be a plausible signature of consciousness.

Conclusion

Determining the sequence of processing stages through which a stimulus passes is an essential goal for cognitive neuroscience. Furthermore, if the GNW theory is correct, assessing whether a brief stimulus reaches a late stage of information processing in which the sensory information is stabilized and is made available to further processors can provide an efficient signature of consciousness. Both of these goals can now be achieved through the use of temporal decoding and of the temporal generalization method. Multivariate decoding of temporal signals provides a sensitive method to probe the time course of information processing. The code is freely available as part of the open-source MNE-Python software (Gramfort et al. 2014; <http://martinos.org/mne/>). Thus, all of the above techniques can now be readily applied to novel problems.

To summarize, our experimental findings suggest that (1) the initial stages of stimulus-evoked brain activity reflect non-conscious processing and are systematically associated with a “diagonal” pattern of temporal generalization; (2) conscious perception relates to a late period of metastability and slow sequential processing, associated with the ignition of a distributed parietal and prefrontal network and with a temporally extended, “square” pattern of temporal generalization. Recently, we have obtained evidence suggesting that these conclusions may generalize to dual-task paradigms such as the attentional blink and the psychological refractory period (Marti et al. 2015). In the future, it will be essential to determine whether they can be validated in additional paradigms.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work’s Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work’s Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L (2009) Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci USA* 106:1672–1677
- Charles L, King J-R, Dehaene S (2014) Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *J Neurosci* 34:1158–1170

- Charles L, Van Opstal F, Marti S, Dehaene S (2013) Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage* 73:80–94. doi:[10.1016/j.neuroimage.2013.01.054](https://doi.org/10.1016/j.neuroimage.2013.01.054)
- Dehaene S, Changeux JP (2011) Experimental and theoretical approaches to conscious processing. *Neuron* 70:200–227
- Dehaene S, Naccache L (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79:1–37
- Dehaene S, Sergent C, Changeux JP (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc Natl Acad Sci USA* 100:8520–8525
- Dehaene S, Changeux JP, Naccache L, Sackur J, Sergent C (2006) Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn Sci* 10:204–211
- Del Cul A, Baillet S, Dehaene S (2007) Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol* 5:e260
- Donders FC (1969) On the speed of mental processes (translation). *Acta Psychol (Amst)* 30:412–431
- Faugeras F, Rohaut B, Weiss N, Bekinschtein TA, Galanaud D, Puybasset L, Bolgert F, Sergent C, Cohen L, Dehaene S, Naccache L (2011) Probing consciousness with event-related potentials in the vegetative state. *Neurology* 77:264–268
- Faugeras F, Rohaut B, Weiss N, Bekinschtein T, Galanaud D, Puybasset L, Bolgert F, Sergent C, Cohen L, Dehaene S, Naccache L (2012) Event related potentials elicited by violations of auditory regularities in patients with impaired consciousness. *Neuropsychologia* 50:403–418
- Friston K (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360:815–836
- Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS (2014) MNE software for processing MEG and EEG data. *Neuroimage* 86:446–460
- Isik L, Meyers EM, Leibo JZ, Poggio T (2014) The dynamics of invariant object recognition in the human visual system. *J Neurophysiol* 111:91–102
- King J-R (2014) Characterizing electro-magnetic signatures of conscious processing in healthy and impaired human brains. Phd Thesis, University of Paris VI, Paris, France
- King J-R, Dehaene S (2014) Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn Sci* 18:203–210
- King JR, Faugeras F, Gramfort A, Schurger A, El Karoui I, Sitt JD, Rohaut B, Wacongne C, Labyt E, Bekinschtein T et al (2013a) Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *Neuroimage* 83:726–738
- King J-R, Sitt JD, Faugeras F, Rohaut B, El Karoui I, Cohen L, Naccache L, Dehaene S (2013b) Information sharing in the brain indexes consciousness in noncommunicative patients. *Curr Biol* 23:1914–1919
- Kouider S, Stahlhut C, Gelskov SV, Barbosa LS, Dutat M, de Gardelle V, Christophe A, Dehaene S, Dehaene-Lambertz G (2013) A neural marker of perceptual consciousness in infants. *Science* 340:376–380
- Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535–540
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
- Marti S, King J-R, Dehaene S (2015) Time-resolved decoding of two processing chains during dual-task interference. *Neuron* 88(6):1297–1307. doi:[10.1016/j.neuron.2015.10.040](https://doi.org/10.1016/j.neuron.2015.10.040)
- Pashler H (1994) Dual-task interference in simple tasks: data and theory. *Psychol Bull* 116:220–244
- Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME (1988) Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 331:585–589
- Posner MI (1978) Chronometric explorations of the mind. Lawrence Erlbaum, Hillsdale, NJ
- Posner MI, Raichle ME (1994) Images of mind. Scientific American Library, New York

- Rangarajan V, Hermes D, Foster BL, Weiner KS, Jacques C, Grill-Spector K, Parvizi J (2014) Electrical stimulation of the left and right human fusiform gyrus causes different effects in conscious face perception. *J Neurosci* 34:12828–12836
- Salti M, Monto S, Charles L, King J-R, Parkkonen L, Dehaene S (2015) Distinct cortical codes and temporal dynamics for conscious and unconscious percepts. *eLife* 4
- Schurger A, Sarigiannidis I, Naccache L, Sitt JD, Dehaene S (2015) Cortical activity is more stable when sensory stimuli are consciously perceived. *Proc Natl Acad Sci USA* 112:E2083–E2092
- Sergent C, Baillet S, Dehaene S (2005) Timing of the brain events underlying access to consciousness during the attentional blink. *Nat Neurosci* 8:1391–1400
- Sigman M, Dehaene S (2005) Parsing a cognitive task: a characterization of the mind's bottleneck. *PLoS Biol* 3:e37
- Sitt JD, King J-R, El Karoui I, Rohaut B, Faugeras F, Gramfort A, Cohen L, Sigman M, Dehaene S, Naccache L (2014) Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* 137:2258–2270
- Sternberg S (1969) The discovery of processing stages: extensions of Donders' method. *Acta Psychol* 30:276–315
- Sternberg S (2001) Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychol (Amst)* 106:147–246
- Strauss M, Sitt JD, King J-R, Elbaz M, Azizi L, Buiatti M, Naccache L, van Wassenhove V, Dehaene S (2015) Disruption of hierarchical predictive coding during sleep. *Proc Natl Acad Sci USA* 122:E1353–E1362
- Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311:670–674
- Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc Natl Acad Sci USA* 108:20754–20759

Sleep and Synaptic Down-Selection

Giulio Tononi and Chiara Cirelli

Abstract Sleep is universal, tightly regulated, and many cognitive functions are impaired if we do not sleep. But why? Why do our brains need to disconnect from the environment for hours every day? We discuss here the synaptic homeostasis hypothesis (SHY), which proposes that sleep is the price the brain pays for plasticity, to consolidate what we already learned, and be ready to learn new things the next day. In brief, new experiments show that the net strength of synapses increases with wake and decreases with sleep. As we discuss, these findings can explain why sleep is necessary for the well-being of neural cells and brain circuits, and how the regulation of synaptic strength may be a universal, essential function of sleep.

Sleep Is Universal and Probably Essential, But Why?

We spend on average a third of our time asleep, but the functions of sleep remain elusive (Mignot 2008; Siegel 2008). This is even more puzzling if one considers that overall, during sleep the brain is almost as active as in waking life: neurons fire at comparable rates as in wake, metabolism is only slightly reduced (Steriade and Hobson 1976). So if sleep is not simply a passive state, during which brain cells can rest, why does it disconnect from the environment, turns on spontaneous activity, experiences vivid dreams but forms no new memories? This question is all the more intriguing since sleep is universal (Cirelli and Tononi 2008). For example, even animals that cannot afford to sleep in the regular manner because they are constantly on the move, such as several species of cetaceans, have found a clever way of cheating with sleep. Thus dolphins continue to swim and breathe with one hemisphere, while the other half of their brain is deep asleep, showing EEG slow waves just as in other mammals (Oleksenko et al. 1992). Though nature offers some

G. Tononi (✉) • C. Cirelli

Department of Psychiatry, University of Wisconsin-Madison, 6001 Research Park Blvd,
Madison, WI 53719, USA

e-mail: gtononi@wisc.edu; ccirelli@wisc.edu

remarkable case studies in sleep self-deprivation—some Alaskan sandpipers briefly forgo regular sleep during the intense mating season (Lesku et al. 2012), and migrating birds make do for a while on a third of their normal sleep (Rattenborg et al. 2004)—there is so far no positive proof of any species that lacks sleep completely (Cirelli and Tononi 2008). We now know that even invertebrates sleep, including fruit flies (Hendricks et al. 2000; Shaw et al. 2000), honeybees (Kaiser 1988), and octopus (Brown et al. 2006), though we still do not know how much spontaneous activity goes on in their sleeping brains. Furthermore, fruit flies not only sleep many hours a night but, just like mammals and birds, if they are kept awake, their sleep becomes longer and more intense—a process called *sleep homeostasis* (Huber et al. 2004b). Therefore, sleep does not seem to be a niche adaptation, like hibernation, which may benefit some species that need to save energy or avoid trouble when they have nothing better to do. Rather, sleep seems to serve some essential, universal function, akin to temperature regulation or digestion (Cirelli and Tononi 2008). What's more, it is a function that requires the brain to work off-line, despite the dangers of being temporarily disconnected from the environment. But what is this function?

Sleep Helps Memory Consolidation, Gist Extraction, and Integration

A popular idea that has gained much attention recently is that sleep may be important for memory. There are now plenty of experiments showing that, after a night of sleep and sometimes just after a nap, newly formed memories are preserved better than if one had spent the same amount of time awake. That is, sleep benefits *memory consolidation* (Rasch and Born 2013). This benefit is especially clear for declarative memories—those one can recollect consciously, such as lists of words or associations between pictures and places. But non-declarative memories, such as perceptual and motor skills, can also profit from sleep. For instance, if you try to reach a target on the screen with the mouse while, unbeknownst to you, the cursor is systematically rotated, you slowly learn to compensate for the rotation and get progressively better. If you sleep over it, you improve further, and your movements become smooth (Huber et al. 2004a). These experimental results fit the common observation that after intensive learning, say practicing a piece over and over on the guitar, performance often becomes fluid only after a night of sleep. It is likely that, when you learn by trial-and-error and repeatedly activate certain brain circuits, many synapses end up strengthening, not only when you play the right notes well, but also when you do it badly, or fumble other notes. The result is that, while by practicing you get better and better on average, your performance remains a bit noisy and variable. After sleep, it is as if the core of what you learned had been preserved, whereas the chaff is eliminated—that is, sleep seems to notch-up the signal-to-noise ratio (Rasch and Born 2013; Tononi and Cirelli 2014). Something

similar may happen also with declarative memories: in the face of the hundreds of thousands of scenes we encounter in waking life, memory is particularly effective at *gist extraction*, where the details (the noise) may be lost, but the main point of what happened to you (the signal) is preserved. So far, it seems that the memory benefits of sleep, especially for declarative memories, are due primarily to non-rapid eye movement (NREM) sleep, but in some instances REM sleep or a combination of NREM-REM cycles may also play a role (Rasch and Born 2013). Of course, while sleep is undoubtedly important for memory consolidation, one should not forget that memories can also consolidate during wake. Moreover, to some extent sleep helps memory consolidation simply because it reduces the interference caused by later memory traces, since when you sleep you stop learning new things (Ellenbogen et al. 2006).

Another process that may benefit from sleep is the *integration of new with old memories*. Psychologists have long recognized that one tends to learn new material better if it has many points of contacts with previous knowledge. For example, a new word in a language you already know fairly well is easier to remember than a new word in a completely unknown language. This process of integration certainly occurs during wake—a memorable stimulus will activate, consciously or subconsciously, a vast network of associations throughout the brain (read synapses), with which it may become linked. However, sleep may be a particularly good time to assess which of the new memories fit better and which worse with the vast amount of organized older memories—also known as schemata—that are stored in brain circuits (Lewis and Durrant 2011). This is because during sleep it is possible to activate a large number of circuits in many different combinations without worrying about the consequences for behavior, something that is not advisable during wake, when one must stick to the situation at hand. For example, in real life it would not be a good idea to reminisce about your father’s old car being of a similar color as that of the large truck that is rapidly approaching. In a dream, instead, it is perfectly fine to put your father in the truck’s driver seat, realize later that it is actually a school bus, and notice that it is filled with old people who resemble your colleagues. Perhaps during sleep your brain is sifting through old memories and trying out which new ones fit best overall, while getting rid of the rest, just as it does with gist extraction.

Sleep Helps Memory, But How? Is It Replay or Synaptic Down-Selection?

The ongoing activity in the brain throughout sleep, then, could have something to do with consolidating memory traces, extracting their gist, and integrating new with old memories (Born et al. 2006; Rasch and Born 2013). This idea is supported by studies performed over the past 20 years, first in rodents and then in primates, which show that patterns of neural activity during sleep often resemble those recorded

during wake (Wilson and McNaughton 1994; Kudrimoti et al. 1999; Nadasdy et al. 1999; Hoffman and McNaughton 2002). For example, when a rat learns to navigate a maze, different hippocampal neurons fire in different places, in specific sequences. Presumably, each sequence is encoded in memory by strengthening the connections between neurons firing one after the other. During subsequent sleep, especially NREM sleep, these sequences are “replayed” above chance (though neither very often nor very faithfully). Based on this evidence, many researchers think that sleep “replay” may consolidate memories by further reinforcing the synaptic connections that had been strengthened in wake, leading to *synaptic consolidation*. There may also be some *system-level consolidation*, based on evidence that over time memories may be shuttled around in the brain. For example, the hippocampus may provide early storage, after which memories are transferred to connected cortical areas, and sleep may help this transfer (Girardeau et al. 2009). However, there is also evidence that the “replay” of neural circuits can also occur in wake, not just in sleep (Karlsson and Frank 2009), and “preplay” can also occur during wake *before* learning (Dragoi and Tonegawa 2011).

An interesting alternative is that sleep may be a time not so much for rehearsal, but for *down-selection* (Tononi and Cirelli 2014). In essence, the idea is this: when the brain sleeps, spontaneous neuronal firing activates many circuits in many different combinations, both new memory traces, which may be particularly prone to reactivation, and old networks of associations. But instead of strengthening whatever synapses are activated the most, which would lead to learning things that never happened, the brain could reverse plasticity rules, and promote the activity-dependent weakening of connections. Indeed, an efficient way to do so would be to implement a selectional, competitive process. For example, synapses that are reactivated most strongly and consistently during sleep would be protected and survive mostly unscathed, whereas synapses that are comparatively less activated would be depressed. This down-selection process would literally ensure the survival of those circuits that are “fittest,” either because they were strengthened repeatedly during wake (the signal, i.e., the right notes on the guitar) or because they are better integrated with previous, older memories (a new word in a known language). Instead, synapses involved in circuits that were only occasionally strengthened during wake (the noise, i.e., fumbled notes on the guitar), or fit less with old memories (a new word in an unknown language), would be depressed and possibly eliminated. In this way, synaptic down-selection during sleep would promote memory consolidation by increasing signal-to-noise ratios, thereby favoring gist extraction and the integration of new memories with established knowledge. As an additional bonus, down-selection would also make room for another cycle of synaptic strengthening during wake. Indeed, there are several indications that sleep, in addition to memory consolidation, gist extraction, and integration, is particularly beneficial to *memory acquisition*: quite a few studies have shown after a night of sleep you can learn new material much better than after having been awake all day.

Finally, down-selection based on the systematic reactivation of neural circuits old and new would also explain why prolonged quiescence and disconnection from

the environment are important—that is, why one needs to be asleep. This is because sleep is the perfect time for the brain to try out many different scenarios without worrying about behaving appropriately in the real world. Only in this way can the brain go through a large repertoire of situations, collect fair statistics about how each synapse is activated in the context of the entire set of stored memories (how well it fits), and reorganize its networks accordingly. Otherwise, the synapses that were strengthened most recently would always be favored (say you spent the entire day trying out the guitar) at the expense of others that are equally important (you also know how to type, and you would not want to forget it), irrespective of how the new memories fit with your previous knowledge.

In the end, cycles of net strengthening of connections during wake followed by net weakening during sleep may constitute an excellent selectional strategy that implements a healthy reality check: neural activity patterns triggered during wake, when the brain is connected with the environment, would tend to be reinforced, whereas activation patterns triggered during sleep, when the brain is disconnected from the environment and makes up its own imaginary scenarios, would be weeded out systematically.

Synapses Get Stronger in Wake and Weaker in Sleep

There is converging evidence for synaptic down-selection during sleep (Tononi and Cirelli 2014). Experiments performed in fruit flies, rodents, and humans, all seem to indicate that the strength of connections among neurons increases during wake and decreases during sleep. For example, when fruit flies spend the day in an environment with plenty of opportunity for interactions with other flies (a “fly mall”), by evening time there are almost 70 % more synaptic spines—the little protrusion where an incoming axon makes contact with a dendrite—than there were in the morning, and this is true throughout their brain. The next morning the number of spines goes back to baseline, but only if flies are allowed to sleep (Bushey et al. 2011). In adolescent mice one sees a similar phenomenon: in the cerebral cortex the number of synaptic spines tends to grow during wake and to decrease during sleep, although the changes are smaller than in flies (Maret et al. 2011; Yang and Gan 2011). In adult rodents it is not the number of synaptic spines that changes with wake and sleep, but their strength. This is indicated by an increase in the number of AMPA receptors in the synapses after wake, and a decrease after sleep (Vyazovskiy et al. 2008). AMPA receptors are responsible for the bulk of excitatory neurotransmission in mammalian brains, and the potentiation or depression of synapses is ultimately achieved by increasing or decreasing their number. Other experiments have shown that, if one stimulates electrically neural fibers in the cortex, the response one gets from the target neurons is larger after a few hours of wake, and smaller after sleep, and we know that these responses are usually larger when synapses are strong, and smaller when they are weak (Vyazovskiy et al. 2008). A similar experiment was performed in humans using transcranial

magnetic stimulation—a short magnetic pulse applied to the scalp to activate the underlying neurons—and high-density EEG to record the strength of the responses of the rest of the cerebral cortex. The results were clear: the longer the subject was awake, the larger the responses, and it took a night of sleep for the responses to return to baseline (Huber et al. 2013).

One should emphasize that exactly how this down-selection process would take place remains unclear, and the account above remains speculative. Indeed, the precise mechanisms are likely to vary in different species, in different brain structures, and in different developmental periods. For example, it is not known whether in invertebrates sleep is accompanied by intense neuronal activation or not—perhaps there the weakening of synapses can be accomplished without having to go through a large repertoire of old memories. Similarly, it may be that NREM sleep is the ideal time for weakening synapses in an activity-dependent manner in the cerebral cortex, due to the occurrence of slow waves; but that in the hippocampus, which does not generate slow waves, down-selection may happen preferentially during the faster, theta waves of REM sleep (Grosmark et al. 2012). Irrespective of the specific mechanisms, the evidence is strong, in several species, that overall synaptic strength goes up during wake and down during sleep. And if this is so, it has implications concerning the role of sleep that go beyond its benefits to memory consolidation and integration, as we will now briefly discuss.

SHY: The Price of Plasticity

Perform a simple experiment: before you go to bed, try and remember as many things as you can that happened to you today. If you are serious and systematic about it, starting with your first thought upon awakening, the first thing you did, what you had for breakfast, where you had breakfast, and so on, the list will be very long, and very boring. Now even this list would be very incomplete. If you were to wear a camera on your head recording all that happened to you, and if we were to then show you snapshots from the recordings, you would suddenly recognize many other things that happened to you that you did not initially recollect. And then there are perceptual and motor skills that you have acquired or refined during the day, such as the guitar piece you practiced. Obviously, over a typical day a lot of things must have left a trace in your brain. We still do not know what proportion of the trillions of synapses in your brain is actually changed by a day of wake: is it 0.01, 1, 10 % or even more? But for sure a lot of synapses must have been strengthened, as suggested not only by your little evening thought experiment, but also by the experimental evidence reviewed in the previous section.

Now the crucial thing to realize is that all this learning, if it is reflected in the strengthening of synapses, does not come for free. First of all, stronger synapses consume more energy. For its weight, the brain is by far the most expensive organ of the body—accounting for almost 20 % of the energy budget—and of that budget, two thirds or more is for supporting synaptic activity. So if we learn by

strengthening synapses, one could say that we wake up with an efficient engine and we end the day with a gas-guzzler. Also, a net strengthening of synapses is a major source of cellular stress, due to the need to synthesize and deliver cellular constituents ranging from mitochondria to synaptic vesicles to various proteins and lipids. Clearly, learning by strengthening synapses cannot go on indefinitely—day after day—and something must be done about it. That something, says the *synaptic homeostasis hypothesis*, also known as SHY, is the down-selection of synapses down to a baseline level that is sustainable both in terms of energy consumption and cellular stress. And that, says SHY, is the essential function of sleep. In short, sleep is the price we pay for being able to learn and adapt to novel environments when we are awake—most generally, it is the price we pay for plasticity. If this is indeed the essential function of sleep, it is only fitting that, as sleep-dependent synaptic down-selection relieves neural cells of the metabolic burdens accumulated during wake in the service of plasticity, it does so in a smart way, all along benefitting memory consolidation and integration, while also resetting the conditions for efficiently acquiring new memories when we wake up. This would not be the first time that evolution catches many birds with one stone.

Acknowledgements Supported by NIMH grant R01MH099231 to CC and GT, and NINDS grant P01NS083514 to CC and GT.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Born J, Rasch B, Gais S (2006) Sleep to remember. *Neuroscientist* 12:410–424
- Brown ER, Piscopo S, De Stefano R, Giuditta A (2006) Brain and behavioural evidence for rest-activity cycles in *Octopus vulgaris*. *Behav Brain Res* 172:355–359
- Bushey D, Tononi G, Cirelli C (2011) Sleep and synaptic homeostasis: structural evidence in *Drosophila*. *Science* 332:1576–1581
- Cirelli C, Tononi G (2008) Is sleep essential? *PLoS Biol* 6:e216
- Dragoi G, Tonegawa S (2011) Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature* 469:397–401
- Ellenbogen JM, Hulbert JC, Stickgold R, Dinges DF, Thompson-Schill SL (2006) Interfering with theories of sleep and memory: sleep, declarative memory, and associative interference. *Curr Biol* 16:1290–1294
- Girardeau G, Benchenane K, Wiener SI, Buzsaki G, Zugaro MB (2009) Selective suppression of hippocampal ripples impairs spatial memory. *Nat Neurosci* 12:1222–1223

- Grosmark AD, Mizuseki K, Pastalkova E, Diba K, Buzsaki G (2012) REM sleep reorganizes hippocampal excitability. *Neuron* 75:1001–1007
- Hendricks JC, Finn SM, Panckeri KA, Chavkin J, Williams JA, Sehgal A, Pack AI (2000) Rest in *Drosophila* is a sleep-like state. *Neuron* 25:129–138
- Hoffman KL, McNaughton BL (2002) Coordinated reactivation of distributed memory traces in primate neocortex. *Science* 297:2070–2073
- Huber R, Ghilardi MF, Massimini M, Tononi G (2004a) Local sleep and learning. *Nature* 430:78–81
- Huber R, Hill SL, Holladay C, Biesiadecki M, Tononi G, Cirelli C (2004b) Sleep homeostasis in *Drosophila melanogaster*. *Sleep* 27:628–639
- Huber R, Maki H, Rosanova M, Casarotto S, Canali P, Casali AG, Tononi G, Massimini M (2013) Human cortical excitability increases with time awake. *Cereb Cortex* 23:332–338
- Kaiser W (1988) Busy bees need rest, too. Behavioral and electromyographic sleep signs in honeybees. *J Comp Physiol A* 163:565–584
- Karlsson MP, Frank LM (2009) Awake replay of remote experiences in the hippocampus. *Nat Neurosci* 12:913–918
- Kudrimoti HS, Barnes CA, McNaughton BL (1999) Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. *J Neurosci* 19:4090–4101
- Lesku JA, Rattenborg NC, Valcu M, Vyssotski AL, Kuhn S, Kuemmeth F, Heidrich W, Kempnaers B (2012) Adaptive sleep loss in polygynous pectoral sandpipers. *Science* 337:1654–1658
- Lewis PA, Durrant SJ (2011) Overlapping memory replay during sleep builds cognitive schemata. *Trends Cogn Sci* 15:343–351
- Maret S, Faraguna U, Nelson A, Cirelli C, Tononi G (2011) Sleep and wake modulate spine turnover in the adolescent mouse cortex. *Nat Neurosci* 14:1418–1420
- Mignot E (2008) Why we sleep: the temporal organization of recovery. *PLoS Biol* 6:e106
- Nadasdy Z, Hirase H, Czurko A, Csicsvari J, Buzsaki G (1999) Replay and time compression of recurring spike sequences in the hippocampus. *J Neurosci* 19:9497–9507
- Oleksenko AI, Mukhametov LM, Polyakova IG, Supin AY, Kovalzon VM (1992) Unihemispheric sleep deprivation in bottlenose dolphins. *J Sleep Res* 1:40–44
- Rasch B, Born J (2013) About sleep's role in memory. *Physiol Rev* 93:681–766
- Rattenborg NC, Mandt BH, Obermeyer WH, Winsauer PJ, Huber R, Wikelski M, Benca RM (2004) Migratory sleeplessness in the white-crowned sparrow (*Zonotrichia leucophrys gambelii*). *PLoS Biol* 2:E212
- Shaw PJ, Cirelli C, Greenspan RJ, Tononi G (2000) Correlates of sleep and waking in *Drosophila melanogaster*. *Science* 287:1834–1837
- Siegel JM (2008) Do all animals sleep? *Trends Neurosci* 31:208–213
- Steriade M, Hobson J (1976) Neuronal activity during the sleep-waking cycle. *Prog Neurobiol* 6:155–376
- Tononi G, Cirelli C (2014) Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration. *Neuron* 81:12–34
- Vyazovskiy VV, Cirelli C, Pfister-Genskow M, Faraguna U, Tononi G (2008) Molecular and electrophysiological evidence for net synaptic potentiation in wake and depression in sleep. *Nat Neurosci* 11:200–208
- Wilson MA, McNaughton BL (1994) Reactivation of hippocampal ensemble memories during sleep. *Science* 265:676–679
- Yang G, Gan WB (2011) Sleep contributes to dendritic spine formation and elimination in the developing mouse somatosensory cortex. *Dev Neurobiol* 72:1391–1398

Psyche, Signals and Systems

Costas A. Anastassiou and Adam S. Shai

Abstract For a century or so, the multidisciplinary nature of neuroscience has left the field fractured into distinct areas of research. In particular, the subjects of consciousness and perception present unique challenges in the attempt to build a unifying understanding bridging between the micro-, meso-, and macro-scales of the brain and psychology. This chapter outlines an integrated view of the neurophysiological systems, psychophysical signals, and theoretical considerations related to consciousness. First, we review the signals that correlate to consciousness during psychophysics experiments. We then review the underlying neural mechanisms giving rise to these signals. Finally, we discuss the computational and theoretical functions of such neural mechanisms, and begin to outline means in which these are related to ongoing theoretical research.

Introduction

It was with considerable surprise that, 30 years later, in examining the literature of modern psychology I found that the particular problem with which I had been concerned had remained pretty much in the same state in which it had been when it first occupied me. It seems, if this is not too presumptuous for an outsider to suggest, as if this neglect of one of the basic problems of psychology were the result of the prevalence during this period of an all too exclusively empirical approach and of an excessive contempt for ‘speculation’. It seems almost as if ‘speculation’ (which, be it remembered, is merely another word for thinking) had become so discredited among psychologists that it has to be done by outsiders who

Costas A. Anastassiou and Adam S. Shai are equal contributors.

C.A. Anastassiou (✉)
Allen Institute, Seattle, WA 98109, USA
e-mail: asymptotics@googlemail.com

A.S. Shai
Allen Institute, Seattle, WA 98109, USA

Department of Bioengineering, California Institute of Technology, Pasadena, CA 91125, USA

Department of Biology, Stanford University, Palo Alto, CA, USA

have no professional reputation to lose. But the fear of following out complex processes of thought, far from having made discussion more precise, appears to have created a situation in which all sorts of obscure concepts, such as ‘representative processes’, ‘perceptual organization’, or ‘organized field’, are used as if they described definite facts, while actually they stand for somewhat vague theories whose exact content requires to be made clear. Nor has the concentration on those facts which were most readily accessible to observation always meant that attention was directed to what is most important. Neither the earlier exclusive emphasis on peripheral responses, nor the more recent concentration on macroscopic or mass processes accessible to anatomical or electrical analysis, have been entirely beneficial to the understanding of the fundamental problems.

– Friedrich Hayek, Preface to *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology* (1953).

In 1920, a 21-year-old Friedrich Hayek (later to become the famous economist and winner of the 1974 Nobel Prize in Economic Sciences) wrote one of the first explicit proposals linking the coordinated activity of neural assemblies to consciousness and the representation of percepts in the brain (Hayek 1991). Though Hayek would devote the majority of his adult life to economic theory,¹ he would, some three decades later in 1953, publish an extended book on those same ideas in *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology* (Hayek 1999).² The general “problem of theoretical psychology” that Hayek introduced in *The Sensory Order* was to first describe what, and then explain how, physical states of the brain give rise to sensory perception. To satisfy these criteria he postulated a mechanism for how the collective action of individual neurons could carry out a highly complex hierarchical classification function and how such aggregate activity binds sensory primitives to represent percepts—a defining problem still fundamental to modern neuroscience. By recasting the problem of perceptual representation in terms of classification, Hayek made a great leap forward in suggesting a specific framework of neural processing that accounts for our subjective experience. The mechanistic descriptions offered by Hayek point to unparalleled insightfulness at the conceptual level, ultimately bridging the gap between the seemingly ineffable psyche and the algorithmic framework of computation.

Theoretical (and often philosophical) work has continued in the decades since Hayek’s work, but perhaps the most progress has been in identifying biophysical signals that correlate to different behavioral and psychological states. Most typically, electrical activity, as measured via electroencephalography (EEG) or

¹There has been some discussion about the relationship between his thought in theoretical psychology and economics, especially as it relates to the distribution of information in complex networks of individual nodes, e.g., neurons in the brain or humans in a society (Butos and Koppl 2007; Caldwell 2004; Horwitz 2000).

²Interestingly, Hayek considered this work to be one of his most important intellectual achievements and was disappointed that it did not achieve the popularity of his others works (Caldwell 2004).

fluctuations of magnetism assayed via magnetoencephalography (MEG) gathered from the scalp of humans, has been shown to correlate with behavioral and psychological states. An offspring of such studies is the well-known framework of the neural correlates of consciousness (or NCC), i.e., the minimal set of neural events and mechanisms jointly sufficient for a specific conscious percept. The NCC framework, first proposed by one of the discoverers of DNA structure and Nobel prize winner, Francis Crick, and his colleague Christof Koch, was suggested as a scientific framework in which to study consciousness (Crick and Koch 1990, 2003). Generally, the study of consciousness can be separated into studying “contents” and “level.” The contents of consciousness refer to those perceptual objects that a subject is aware of, for instance, when a subject reports being aware of a tree in their visual field. Level, on the other hand, refers to the continuum spanning from dreamless sleep to normal waking life.

The use of NCC, studying both contents and level, has yielded a fruitful but extremely nuanced list of candidate signals that correlate (in varying degrees and with varying evidence) with consciousness and other related subjects, like attention and decision-making. Due to the necessary use of noninvasive techniques in humans, these signals are often found using EEG or imaging techniques such as functional magnetic resonance imaging (fMRI). Alternatively, in a clinical setting, human patients that have to undergo brain surgery (e.g., to treat epilepsy) live days with intracranial depth electrodes implanted in their brains recording extracellular voltage time-series, allowing neuroscientists to work with them and study how cognitive processing is related to neural signals. Thus, when measured with EEG, MEG, or depth electrodes, the NCC usually consist of modulations in amplitude of these extracellular signals (alongside their timing) or modulations of oscillatory power in certain frequency bands. When measured with fMRI, blood-oxygen level-dependent (BOLD) signals are used as a proxy for neural activity and to find spatial locations of activation mainly in the primate brain. Despite the immense advances in this kind of research, they have taken place largely independent from more theoretical concerns, like those discussed by Hayek.

To understand psychological phenomena, neuroscience must find mechanistic explanations for how these signals reflect, support or even constitute the conscious mind, ultimately explaining theoretical concerns through an understanding of the function of neurons and the circuits they compose. Moreover, we will show how an investigation of the details of physiology and anatomy of the brain can drive the creation of experimentally testable psychological theories. Importantly, neuroscience is now at a point where biophysical and anatomical details can be used to close the gap between experimental neuroscience, psychology and theoretical concerns. In this chapter we introduce and discuss the tight relationship between abstract theoretical concerns, detailed physiological and anatomical facts, and population signals often used in psychophysics experiments. Although much work has been done to find explanations that relate signals to psychological phenomena, it is important to realize that it is the physiology and anatomy of neurons and the networks they create that actually compute and perform tasks in the brain. In other words, within the framework presented herein, the neural substrate of

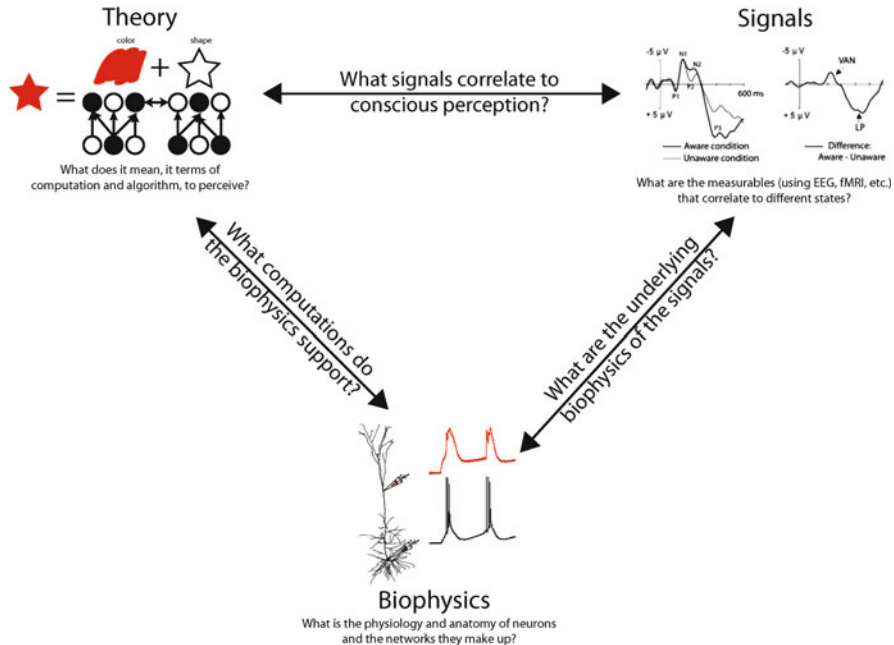


Fig. 1 Signals correlated to conscious perception and theoretical concerns can be connected by considering the biophysics of signals and the computations they perform. Theory concerns itself with what it means in terms of computation and algorithm to consciously perceive something. Signals refer to the population level measurements found in the psychophysics literature (e.g., EEG, fMRI, ECoG). The underlying biophysics of these signals can be uncovered using the tools of experimental neuroscience, and then the computational functionalities of networks made from those biophysics can be explored to bridge theory and signals

psychology is cells and their networks and not (directly) extracellular fields, oxygenation levels, or frequencies in certain bandwidths (though alternative ideas exist; Hameroff 1994; McFadden 2002; Pockett 2012). Thus, theories of consciousness and perception acknowledge that the signals mentioned are proxies for the activity of cells and their networks. The method is thus easily described by a triumvirate of areas of study (in no particular order) related to each other as shown in Fig. 1. We will quickly introduce these three concepts and then delve into them more concretely in the subsequent sections of this chapter.

First are the empirically reported signals that correlate with psychological phenomena. As discussed, these can include signatures of the EEG, anatomical locations found via fMRI, extracellular recorded spiking of cells in the thalamocortical system, and power spectrum analysis in different bands. Second are the theoretical considerations regarding psychological phenomena. These include questions regarding computational and functional concerns; for example, what does it mean in terms of a general algorithm to attend to something or represent a conscious percept? Answers to these questions are often given using some mathematical framework, for instance Bayesian inference (Knill and Pouget

2004; Ma et al. 2006; Yuille and Kersten 2006), predictive coding (Carandini and Ringach 1997; Rao and Ballard 1999), integrated information theory (Oizumi et al. 2014), or the free-energy principle (Friston 2010), or they can take a more conceptual form such as neural Darwinism (Edelman 1993), global workspace theory (Baars 2005), or indeed the ideas of Hayek and their modern extensions like the cognit (Fuster 2003, 2006).

Bridging the empirical signals and theoretical concerns are the biophysical mechanisms. One natural area of study arises in elucidating the physiological underpinnings of signals that correlate to specific psychological states. For instance, given a specific EEG amplitude occurring over the visual cortex, which networks, cell types, transmembrane currents, etc., contribute to that signal? Because these anatomical and physiological details are the substrates of neural computation, we can then delve into the computational role these physical mechanisms play. These questions connect high-level (macro-scale) theory, low-level (micro-scale) biophysical details, and mid-level (meso-scale) psychophysical signals.

In this chapter we explore how distinct biophysical processes connect between signals and psyche. Specifically, using the physiology and anatomy of pyramidal neurons in the neocortex, we explore a mechanism for perceptual binding. Notably, we focus exclusively on the contents of conscious perception. It is important to state at the onset that the connections presented herein are just one of a set of plausible frameworks for understanding how the different scales studied by neuroscientists connect to each other. This chapter is meant not to present the final word on how to comprehensively think about the micro-, meso-, and macro-scales in neuroscience as they relate to consciousness but, instead, to present, by way of example, one possible path to bridge these multiple concerns. Importantly, the task of finding the relationship between biophysics, network computation, theory, and psychology is still very much an open area of study.

Signals of Conscious Perception

What processes in the brain accompany and support conscious perception? In the attempt to answer this question, scientists and clinicians have carried out more than a century's work, often under the area of study called psychophysics, to find measurable signals in the brain that correlate to consciousness. In particular, we discuss the evidence for three such neural signatures: (1) late extracellular signals, (2) distributed information sharing in the cortex, and (3) long-range feedback connections within the cortex. As we will see, the boundaries between these topics are often overlapping but have been studied in an independent enough manner to discuss individually (though not necessarily independently). Notably, given that many of these subjects are discussed in other chapters of this book, we review a number of perceptual correlates rather succinctly in order to relate them to the more general framework discussed in the introduction of this chapter.

Late Extracellular Signals

In 1964, Haider et al. (1964) used scalp electrodes to record extracellular signals from humans during a simple detection task. Dim flashes of light were shown to the subjects, who were asked to report perception of these stimuli. When comparing the averaged extracellular signature of seen and unseen trials, a significant difference was found in the amplitude of a negative wave occurring approximately 160 ms after the signal onset, with the amplitude of the negative wave being positively correlated to perception. These visual results were later reproduced in the auditory cortex (Spong et al. 1965).

Similar conclusions were formed in a series of papers in the 1980s and 1990s. Cauller and Kulics performed a go/no-go discrimination task on forepaw stimulation in monkeys (Kulics and Cauller 1986, 1989). They compared the extracellular signal in the somatosensory cortex and found that an early positive component (called P1, occurring about 50 ms after the stimulus) correlated well with the signal strength whereas a later negative component (called N1) correlated with the behavioral report of the signal (interpreted as the conscious perception). In a later study using depth electrodes, the laminar structure of these signals was examined using current source density analysis. Interestingly, the early P1 signal was found to be attributable to a current sink in layer 4, whereas the later N1 signal was attributed to a current sink in layer 1. Later work also showed that the later N1 signal was absent during sleep and anesthesia (Cauller and Kulics 1988).

More recent psychophysical work, using a spectrum of masking techniques, has suggested a variety of different extracellularly recorded signals that might correlate with consciousness. Two of the most plausible seem to be the Visual Awareness Negativity (VAN; Koivisto et al. 2008) and the p3b (also known as p300 or late potential). Discussion of whether these signals correlate with consciousness itself, or with pre- or post-conscious events, is ongoing (for reviews see Koivisto and Revonsuo 2010; Railo et al. 2011). The p3b is a signal occurring in a largely all-or-none fashion from 300 to 400 ms after stimulus onset (Fig. 2a), but it can occur earlier based on expectation (Melloni et al. 2011).³ The VAN (Fig. 2a) shows a more graded response than p3b and occurs from 100 to 200 ms after the stimulus, but it has been shown to occur as late as 400 ms under specific stimulus conditions. One study asked subjects to report the subjective awareness of a change in a visual stimulus. EEG signals in aware and unaware trials from the occipital lobe were compared (Fig. 2a). Both the p3b (referred to as P3 in their figure) and the VAN can be seen to clearly signify the difference in awareness (Koivisto and Revonsuo 2007). We will not review all the differences between these signals and all the evidence for their correlation (or absence of correlation) to conscious perception here, but suffice it to say, there seems to be an NCC in a late signal occurring at least 100 ms after the stimulus onset, extracellularly measurable from the scalp. The

³ Debate over the p3b and what it correlates with has increased recently, with evidence both pointing to (Gaillard et al. 2009; Salti et al. 2015) and against (Silverstein et al. 2015) its status as an NCC.

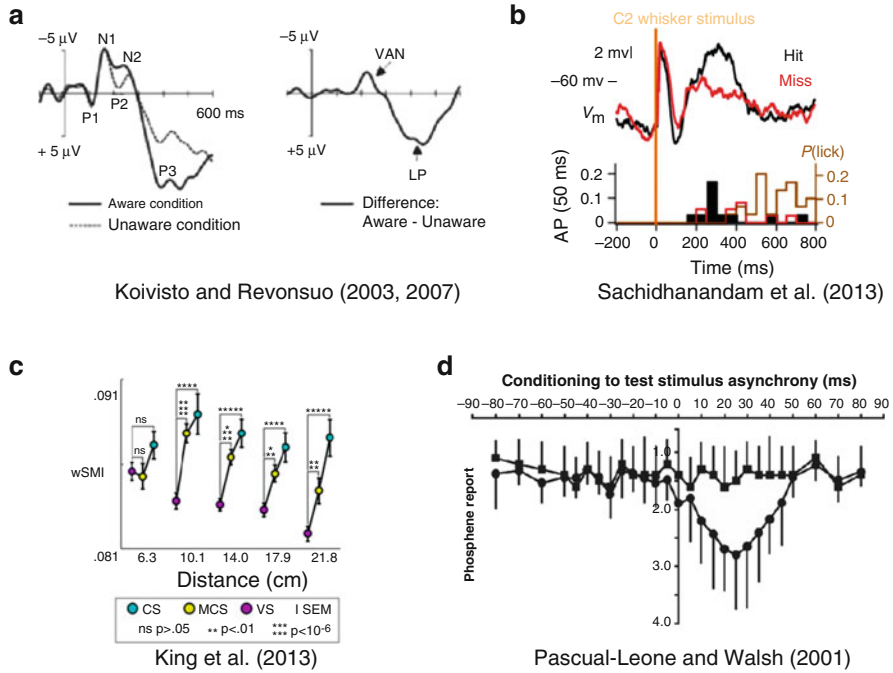


Fig. 2 (a) EEG signals taken from occipital sites during a change blindness task. On the *left* are averaged responses from trials where the subject was aware or unaware of the change. On the *right* is the difference between aware and unaware trials. Data from Koivisto and Revonsuo (2003), figure from Koivisto et al. (2007). (b) The subthreshold membrane potential of a mouse L2/3 pyramidal neuron during a whisker stimulus task. Behavioral hits and misses are shown in *black* and *red*. There are two epochs of depolarization, with the late epoch correlating to the behavioral output. Figure from Sachidhanandam et al. (2013). (c) Weighted symbolic mutual information between EEG sites in control (CS), minimally conscious (MCS), and vegetative (VS) patients. As the distance between sites increases, the differences in wSMI become more and more significant between the different conscious states. Figure from King et al. (2013). (d) Phosphene report after TMS stimulation in area V5 followed by V1, after a time delay shown on the x-axis. When V1 stimulation followed V5 stimulation within ~50 ms, phosphene report was abolished. Figure from Pascual-Leone and Walsh (2001)

VAN is particularly interesting as the timing of this signal corresponds to the timing of the signals measured in the Haider et al. (1964) study as well as the Kulics and Cauller work discussed above.⁴ As argued below, the VAN or p3b might even correspond to recent measurements in behaving rodents.

One of the main advantages of primate experiments is the relatively direct knowledge of what the subjects' perception is, though of course this advantage is offset by more limited access to physiological properties. Rodent experiments have

⁴Though care must be taken not to over-interpret. It is important to realize, for instance, that these signals all come from different perceptual modalities and cognitive tasks.

been used as a model organism for cortical physiology at the synaptic, single-neuron (including dendrites), and small network level. Recent genetic tools (e.g., cre-lines, opsins) have made the mouse a preferred animal in cellular and systems neuroscience, despite the relative difficulty in establishing complex behavioral tasks and inferring perceptual state. By establishing measurable (often population or indirect) signals in primates, experimentalists are now able to find analogous signals in the rodent cortex as they attempt to establish links between behavior and perception. One recent example is from Sachidhanandam et al. (2013) (Fig. 2b). In this experiment, mice were trained to report a whisker stimulus during whole-cell patch recording of single pyramidal neurons in the barrel cortex. Two periods of depolarization were found. The first, occurring within 50 ms of stimulus onset, correlated well with stimulus strength. The second signal, occurring 50–400 ms after stimulus onset, correlated well with the behavioral report. Taking advantage of the animal preparation, optogenetics was used to silence pyramidal neurons during both the early and late epochs. Both types of inhibition abolished the behavioral report. In a control experiment, inactivation of the forepaw somatosensory cortex (and not the whisker cortex) had no effect on performance. These experiments established a causal influence of the late depolarization specifically in the whisker cortex for the perception of whisker deflection.

Taken together, these findings suggest a potential NCC in a late (~150 ms) signal that originates in the upper layers of the neocortex.

Distributed Processing in the Cortex

How distributed is the cortical representation for a given conscious percept? What are the necessary and sufficient conditions related to the communication between different areas of the brain and representation of such percepts? Here we review the evidence pointing to the distributed nature of cortical percepts.

Perhaps the earliest work hinting at the distributed mode in which the cortex operates was given by the pioneering physiologist Flourens, who sought to test the theory of localized function in the brain made popular by phrenologists like Gall and Spurzheim around the turn of the nineteenth century.⁵ Flourens removed different parts of the brain in rabbits and pigeons and assessed a range of behavioral abilities. Although he was able to ascribe differences in function between the cerebellum and cerebrum, for instance, he was unable to relate different parts of the cerebrum to different cognitive and memory-dependent behaviors, ultimately positing that memory and cognition were highly distributed throughout the cerebrum (Flourens 1842).

⁵This task was actually assigned to Flourens by the French Academy of Sciences in Paris, on order of Napoleon Bonaparte. Gall was not seen to have carried out his experiments with ample scientific rigor by the Academy (Pearce 2009).

Alongside medical results from the injured soldiers of WWI (Goldstein 1942) and a number of famous case studies (Harlow 1999), this line of study was continued a century later by Lashley. In this body of work (Lashley 1929, 1950), Lashley aimed to study the relationship between cerebral damage and cognitive behavior, wanting to more quantitatively explain results in human patients with cortical damage who had their visual discrimination assessed by using more invasive experiments in rodents, very similar to those of Flourens. In this work, rats were trained to run through a maze. Upon removing varying volumes of cortex in different areas, rats were reintroduced into the maze, and their ability to complete the maze was assessed. Lashley found that the maze-running ability was related to the volume, but importantly not the location, of the cortical lesion. He thus posited that the ability to run through the maze was not contained in any specific local part of the cerebrum but was, instead, distributed among the entirety of the cortex.

One caveat of the work presented so far is that it is often not explicitly testing the distributed nature of a conscious percept per se but instead a more general cortex-dependent behavior. More recently, psychophysical experiments in humans have suggested that widely distributed cortical activity is associated with conscious perception, whereas activity more localized to the primary sensory areas is not. Using intracortical EEG, Gaillard et al. (2009) used a masking paradigm to compare conscious and unconscious extracellular signatures. They found that conscious perception of the stimulus was associated with widely distributed voltage deflections sustained across the cortex, increased beta (12–20 Hz) synchrony across the cortex, as well as gamma (30–60 Hz) power. The timing of these changes was late, occurring most obviously 300 ms after stimulus presentation (this was interpreted as being the p3b, though significant differences could be measured starting at 200 ms). Other similar studies showed that more localized gamma band activity relegated to the visual cortex accompanied conscious perception (Fisch et al. 2009), though follow-up studies argued that these signals were related more to pre- or post-conscious processing (e.g., decision making and report; Aru et al. 2012) than with conscious perception itself, a general weakness of the contrastive method (Aru et al. 2012; de Graaf et al. 2012; Tsuchiya et al. 2015).

Two recent studies used mathematical concepts related to information sharing across the cortex to successfully quantify the amount of consciousness in patients. King et al. (2013) used weighted symbolic mutual information, a novel measure of information sharing, between pairs of EEG recording sites (Fig. 2c). Importantly, in comparing this information measure using different distances between electrodes, it was found that differences between different levels of consciousness (e.g., vegetative vs. minimally conscious vs. healthy) were most significant for mid- to long-range distances, implicating information sharing between far-away parts of cortex in consciousness. Casali et al. (2013) used TMS evoked potentials to assess the amount of integration and differentiation distributed across the scalp EEG of patients. Importantly, this method was able to accurately and quantifiably assess the level of consciousness in patients undergoing anesthesia, sleep (Massimini et al. 2005), and varying degrees of brain injury. Similar results were more recently shown by Sarasso et al. (2015) by comparing propofol and xenon anesthesia, which

induce states of unconsciousness with no dreams, to dream-inducing ketamine anesthesia. In propofol and xenon anesthesia, integration and differentiation measures were found to be low, whereas in ketamine, these same measures were high. These two studies show that the concept of long-range distributed information sharing is not only a qualitatively useful correlate of consciousness but is also quantifiable and workable in a medically applicable setting. Similar studies using transfer entropy measures have been used to study anesthesia in rats (Imas et al. 2005).

How distributed the representation for a conscious percept needs to be is a matter of ongoing debate. For visual perception, it is quite clear that V1 is generally necessary but not in itself sufficient to support a conscious content (Blake and Fox 1974; Crick and Koch 1995; Cumming and Parker 1997; Gawne and Martin 2000; Rees et al. 2002), though it is unclear if information processing needs to reach extrastriate areas or the most frontal regions or the entirety of cortex. Whatever the case, long-range communication in the cortex⁶ between at least several centimeters in a human (or on the order of a millimeter in the mouse) is a necessary condition for representation of a conscious percept.

Feedback Processing

A separable but not completely independent area of study from the distributed nature of processing in the cortex is the study of feedback processing of extrastriate areas or frontal regions to primary visual cortex. Here, the data in any one study do not often explicitly implicate feedback processing but are instead interpreted to be feedback from considerations like timing and anatomy.

The timing of extracellularly measured potentials that correlate to consciousness, like the VAN discussed previously, suggests that they might have their origin in long-range feedback connections from other areas of cortex. The sensory driven, feedforward step of information processing follows a stereotyped succession of cortical areas and is completed in ~100 ms (Lamme and Roelfsema 2000). Indeed, many theories of consciousness rest on this fact, and some even go so far as to equate recurrent processing with consciousness (Lamme 2006). Experiments using TMS and other stimulation techniques have tested the causal influence of late, presumably long-range feedback processing, on perception. Multiple studies using different sensory paradigms have now shown interruption of perception by TMS over V1 during two distinct time periods, the early one interpreted to be the feedforward sweep and a later one (>200 ms) interpreted to be a feedback sweep

⁶One interesting possibility is that such long-range communication is mediated through the thalamus via L5b pyramidal neurons and not directly within the cortex. Some evidence exists that such a pathway is indeed the main mode in which different areas of cortex communicate with each other (Sherman and Guillery 2002, 2011).

(Heinen et al. 2005; Juan and Walsh 2003). Additionally, phosphenes induced by TMS over V5 (an extrastriate visual area) can be reduced by a lagging TMS pulse over V1, presumably interrupting the feedback of information from V5 to V1 (Fig. 2d; Pascual-Leone and Walsh 2001).

Another line of evidence comes from single cell recordings, showing that cells in the cortex continue spiking past initial feedforward activity. Many cells in macaque V1 have been found to possess dynamic orientation tuning, having precise tuning to one orientation starting at around 50 ms and then inverting at 120 ms (Ringach et al. 1997). Network simulations have shown that feedback, but not feedforward, networks can recapitulate these dynamic tuning curves (Carandini and Ringach 1997). Furthermore, single unit recordings have shown the early firing of cells codes tuned for the general category (e.g., face), whereas later spiking, ~165 ms, was tuned for specific identity (Sugase et al. 1999). Finally, inactivation of higher areas of cortex (e.g., area MT) greatly altered the response properties of cells in lower areas (e.g., V1 and V2), where feedback axons project (Nowak and Bullier 1997).

A host of studies using a technique called backwards masking might also be explained by the need for feedback processing in consciousness. In backwards masking, a target stimulus is followed, after ~50 ms, by a mask (Breitmeyer and Ogmen 2000). The subject is not aware of the target stimulus, even though on trials without a mask the target is consciously perceived. One explanation for this phenomenon is that, while the feedforward information flow through the cortex is preserved, the feedback signals conflict with the mask, rendering the target unconscious. A similar effect is found in patients with V1 lesions. These so-called “blindsight” patients retain the ability to perform forced choice tasks even though they can no longer consciously perceive visual stimuli into the affected visual field (Weiskrantz 1986). Although the exact neural underpinnings of blindsight are unknown, one candidate mechanism implicates the largely intact feedforward sweep in the retained information processing capabilities and the disturbed feedback processing in the absence of consciousness (Lamme 2001). Feedback processing has also been implicated in “contextual modulation,” which is the altering of cellular responses by changes of the stimuli outside of their classical receptive field. Interestingly, blindsight of stimulus that would normally create contextual modulation abolishes such modulation (Zipser et al. 1996), as does anesthesia (Lamme et al. 1998).

Biophysical Foundations of Signals Associated with Conscious Perception

The aforementioned relationships between conscious perception and a number of characteristic signals and signatures point to the importance of understanding the neural substrate of these signals. Such understanding bridges the gap between the

underlying cellular biophysics, the network effects, and the high-level behavioral readouts. To gain insights into the signals associated with conscious perception, it is important to understand the underlying physics, in terms of the physical laws governing the generation of these signals as well as the neural origins that brings them about.

We first present the physics underlying electric measurements in the brain ('Biophysics Related to Electric Measurements'). We have chosen to specifically focus on electric signals and measurements such as the VAN as they have produced the largest body of evidence in terms of psychophysics of conscious perception. (Later in this chapter we also present other methods that have impacted or will potentially critically impact the field.) In a next step, we introduce the most significant cellular contributors of electric activity in brain matter as a means to understand which processes (synapses, cells, circuits, etc.) contribute to these signals ('Biological Electric Field Contributors'). Finally, we present the most prominent methods and technologies used to monitor brain activity ('Monitoring Neural Activity').

The previous section featured results using several different types of electrical measurements, including EEG (Koivisto and Revonsuo 2010), single unit recordings (Sugase et al. 1999), and depth electrodes to compute the power of different frequency spectrum (Aru et al. 2012), as well as both local field potential (LFP) and current source density (CSD) recordings (Kulics and Cautler 1986). These techniques as well as others used in the field of neuroscience will be presented. Additionally, the biophysical underpinnings of the late current sink in layer 1 (Kulics and Cautler 1986) that correlates to conscious perception is discussed.

Biophysics Related to Electric Measurements

Charge transfer across the membrane of all structures in brain matter such as neurons, glial cells, etc., induces so-called extracellular sinks and sources that, in turn, give rise to an extracellular field, i.e., a negative spatial gradient of the extracellular voltage (V_e) measured in comparison to a distant reference signal. The physics governing such events are described by Maxwell's equations. In their simplest form, Maxwell's equations of electromagnetism dictate that V_e depends on the transmembrane current amplitude (I), the conductivity of the extracellular medium (σ) and the distance between the location of the ionic flux and the recording. Specifically, when assuming a so-called point-source (i.e., when a localized current injection occurs within an electrically conductive medium), the relationship between the aforementioned variables and the resulting V_e is (Fig. 3a):

$$V_e(d) = \frac{I}{4\pi\sigma d}$$

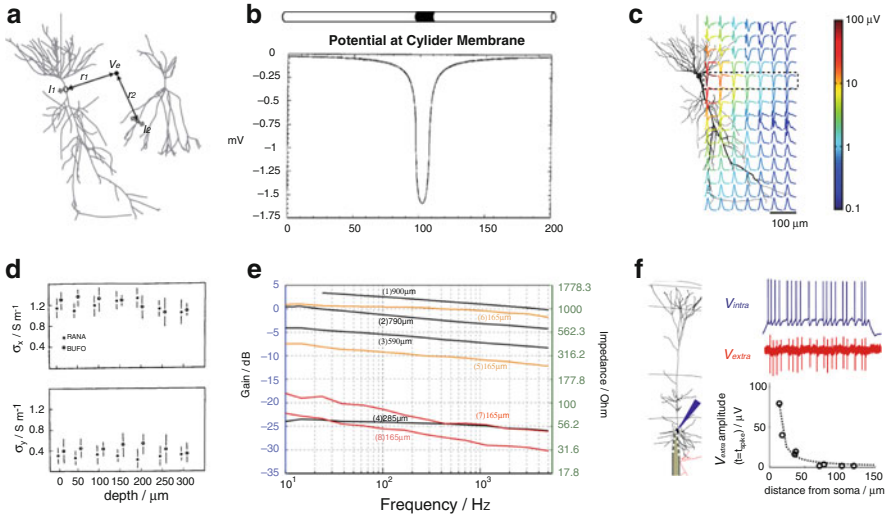


Fig. 3 Biophysics of extracellular signatures and conductivity of the extracellular medium. (a) Illustration of V_e calculation in a population through the superposition of contributions from all compartments in all cells. Individual compartment contributions are primarily determined by their transmembrane currents and distances from the electrode. (b) Charge transfer elicited across the membrane (dark region) of a long, small diameter cable gives rise to an extracellular field. The extracellular potential close to the cable was calculated using the line-source and the cylinder-source approximation. The difference between the two approximations is very small (they overlap). (c) Simulated location dependence of the extracellular action potential (EAP) waveform of a pyramidal neuron. The peak-to-peak voltage range is indicated in this simulation by the color of each trace. EAPs are calculated at the location of the start of each trace. EAP amplitude decreases rapidly with distance. (d) Experimentally obtained values of components of the conductivity tensor in the frog (*Rana*) and toad (*Bufo*) cerebellum as a function of depth. (e) In vivo measurements of impedance as a function of cortical depth in monkey. (f) Microscopic measurements of the relationship between intracellular and extracellular spike signals in rodent slice. Whole-cell patched neurons are brought to spike (blue line) and a proximally positioned extracellular silicon probe with eight contacts is used to record the elicited extracellular voltage transients (red line). At the initiation time of the spike, the extracellular negativities (red) associated with the intracellular spikes attenuate with distance from the soma (see also panel c), with the attenuation occurring per the point-source approximation. Figure contributions are from (a, c) Schomburg et al. (2012), (b) Holt and Koch (1999), (d) Nicholson and Freeman (1975), (e) Logothetis et al. (2007), (f) Anastassiou et al. (2015)

Based on the point-source equation, one can note the following: first, there is an inverse relationship between distance d and the amplitude of the resulting voltage deflection V_e , i.e., the farther away to recording site is from the location of the current point-source, the larger the attenuation of the amplitude of the V_e -deflection; the stronger the point-source I , the larger the V_e -deflection; finally, the conductivity of the extracellular medium critically impact propagation of the signals from the point-source to the recording site.

Notably, when the source is not limited to a point but instead possesses physical extent, the approximation needs to be re-formulated accordingly to account for such

physical extent. For example, when charge transfer takes place along the elongated, cable-like morphologies of neurons, it gives rise to a spatially distributed extracellular source not compatible with the aforementioned point-source expression. Probably the most prominent such approximation accounts for the field induced by a linear, one-dimensional (line) source of infinitesimally small diameter. The line source approximation (LSA) makes the simplification of locating the transmembrane net current for each neurite on a line down the center of the neurite. By assuming a line distribution of current, V_e is described via a two-dimensional solution in cylindrical coordinates. For an elongated current source of length Δs , the resulting $V_e(r, q)$ is given by:

$$V_e(r, q) = \frac{1}{4\pi\sigma} \int_{-\Delta s}^0 \frac{I}{\Delta s \sqrt{r^2 + (q-s)^2}} ds = \frac{I}{4\pi\sigma\Delta s} \log \left(\frac{\sqrt{q^2 + r^2} - q}{\sqrt{l^2 + r^2} - l} \right)$$

where r is the radial distance from the line, q the longitudinal distance from the end of the line, and $l = \Delta s + q$ is the distance from the origin of the line. Holt and Koch (1999) analyzed the accuracy of the LSA and found it to be highly accurate except at very close distances (i.e., about 1 μm) to the cable (see also Rosenfalck 1969; Trayanova and Henriquez 1991; Fig. 3b). The LSA has been the primary method of calculating extracellular voltages arising from transmembrane currents (Gold et al. 2006, 2009; Holt 1998; Holt and Koch 1999; Pettersen and Einevoll 2008; Fig. 3c).

Notably, the aforementioned relationships assume that the extracellular medium in the brain is described via electrostatics and not by much more elaborate elements of electrodynamics. Furthermore, a widespread assumption is that the extracellular medium is isotropic and homogeneous. What evidence exists for such claims to be made? It turns out that this question has remained unresolved, with a number of studies reporting an anisotropic and homogeneous σ (Nicholson and Freeman 1975; Logothetis et al. 2007) (Fig. 3d, e) to strongly anisotropic and inhomogeneous (Goto et al. 2010; Hoeltzell and Dykes 1979; Ranck 1973) and, finally, even of capacitive nature (Bédard and Destexhe 2009; Bédard et al. 2004; Gabriel et al. 1996).

Part of the difficulty in determining the properties of σ , especially at the local, microscopic scale, has to do with the inhomogeneity of the brain as a structure. In that sense, the questions to be answered are where, in what species, in what frequency band and at what spatial scale should σ be measured. The danger is that measuring σ over larger volumes leads to possibly quite different results (attributed to averaging) than recording σ over tens of μm . Moreover, measuring σ within distances of tens micrometers, i.e., the relevant spatial scale for signals related to spiking, poses significant technical challenges given the large number of sites (both for current injection and voltage recording) that need to be positioned within μm -distances and the resulting tissue deformation/damage.

Recently, detailed whole-cell patch recordings of excitatory and inhibitory neurons in rat somatosensory cortex slices were performed in parallel to positioning

a silicon probe in the vicinity of the patched somata, allowing concurrent recording of intra- and extracellular voltages (Anastassiou et al. 2015). Using this experimental setup, the authors characterized biophysical events and properties (intracellular spiking, extracellular resistivity, temporal jitter, etc.) related to extracellular spike recordings at the single-neuron level. It was shown that the extracellular action potential (EAP) amplitude decayed as the inverse of distance between the soma and the recording electrode at the time of spike (Fig. 3f). The spatial decay of the EAP-amplitude at the spike time was very close to the prediction of the point-source approximation: at the spike time, transmembrane charge transfer was still spatially localized (close or at the axon initial segment), resulting effectively in a point-source. Even fractions of a ms after the spike time, the relationship between the EAP-amplitude and distance was shown to become more intricate as more extended sections of the cellular morphology acted as sources, leading to more complex superposition rules (e.g., based on the LSA). On that limit, various contributions of a cell's different compartments need to be accounted for. Interestingly, in the same experiments, a time lag was observed at the extracellular spike waveform with increasing distance of the electrode location from the cell body with respect to the spike time at the soma. While such time lags could be explained by the presence of a non-ohmic extracellular medium, the authors showed that they were actually attributed to the spatial propagation of the action potential along the neural morphology, i.e., backpropagating action potentials. Finally, this study demonstrated that different cortical layers exhibited different conductivity, with the conductivity of layer 4 being higher than the conductivity of layer 2/3 and 5, i.e., an observation in line with the finding that layer 4 possesses a higher density of neurons compared to layers 2/3 and 5.

Do these observations hold *in vivo*? A number of experimental studies have appeared offering compelling insights into the physics of the extracellular medium. Nicholson and Freeman (1975) studied the conductivity profile in the cerebellum of bullfrogs using current injections through micropipettes and concluded that it is anisotropic, homogeneous, and purely ohmic, with later measurements by Logothetis et al. (2007) confirming these observations (Fig. 3d, e). Yet, others found the extracellular medium to be strongly anisotropic and inhomogeneous (Hoeltzell and Dykes 1979; Ranck 1973) or even of capacitive nature (Gabriel et al. 1996; Bédard et al. 2004; Bédard and Destexhe 2009). In a more recent study, Goto et al. (2010) used extracellular recordings to measure the conductivity profile along the entire somatosensory barrel cortex in rodents using depth multi-electrode recordings and reported that radial and tangential conductivity values varied consistently across the six neocortical laminae. Thus, they showed that the electric properties of the extracellular medium in the living animal were anisotropic and inherently inhomogeneous, agreeing with the *in vitro* findings of Anastassiou et al. (2015). Importantly, in their work Goto and colleagues provided evidence that (at least for frequencies less than 500 Hz) σ can be assumed to be purely ohmic. Based on the aforementioned, the temporal characteristics of the extracellular field and signals like the VAN are not due to extracellular medium properties but, instead, solely attributed to cellular functioning.

Biological Electric Field Contributors

Given the aforementioned biophysics dictating how transmembrane currents are generated and propagated in brain matter, what cellular processing gives rise to these electric signals? Here we present the most important contributors of the extracellular field. In principle, any charge transfer along the membranes of the neural morphology elicits extracellular sinks and sources, as will be discussed below. (For a more thorough treatise, the interested reader is pointed to Buzsáki et al. 2012; Einevoll et al. 2013.)

Synaptic Activity

In physiological situations, synaptic activity and **postsynaptic currents**, in particular, are often the most prominent sources of extracellular current flow. While the majority of individual synaptic connections induce fairly small extracellular signals (e.g., Bazélot et al. 2010; Glickfeld et al. 2009), thousands of synapses are present along a single neuron's morphology (e.g., a rat layer five pyramidal neurons has approximately 10,000 synapses along its processes). Thus, even if the individual contribution of such postsynaptic events is fairly small, the fact that thousands of them may become co-activated within a small time increment suggests a substantial overall effect (Fig. 4a). Furthermore, the time constant of synaptic events can vary substantially: while the time constant of fast excitatory AMPA- and inhibitory GABA subtype A-receptors ranges approximately 1–15 ms (Hille 1992), excitatory NMDA and inhibitory GABA subtype B-receptor dynamics can be particularly slow (i.e., 50–300 ms; Pérez-Garci et al. 2006) and, as such, may readily contribute to slow bands of the electric signal (Elul 1971; Logothetis and Wandell 2004).

The influx of cations when excitatory synaptic input impinges along the neural membrane from the extracellular into the intracellular space gives rise to a local **extracellular sink**. To achieve effective electroneutrality within the time constants of relevance for systems neuroscience, the extracellular sink needs to be balanced by an **extracellular source**, that is, an opposing ionic flux from the intracellular to the extracellular space, along the neuron. In this case, the counter-flux is termed **passive or return current**. It follows that such passive return currents do not only depend on where synaptic input impinges along the neural morphology but also on the actual morphological features of the neuron itself. For example, impinging inputs in one area of the elongated morphology of pyramidal neurons gives rise to passive return currents along the same neuron (Fig. 4a). On the other hand, the symmetric location of the dendrites of inhibitory basket cells does not allow the formation of such strong passive return currents due to cancellation effects, even when these neurons receive strong synaptic input (Pettersen and Einevoll 2008; Reimann et al. 2013). Depending on the location of the sink current(s) and its distance from the source current(s), a dipole or a higher-order n -pole is formed.

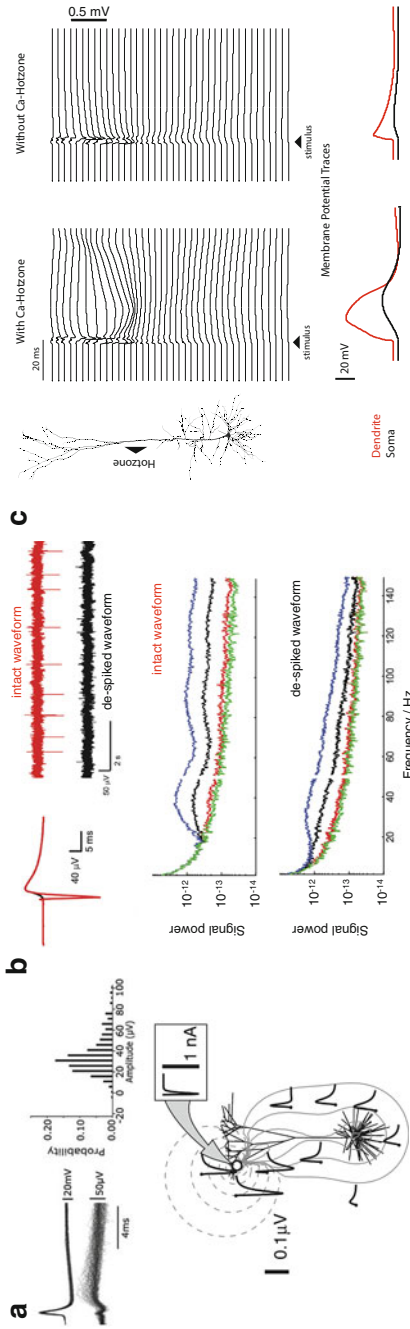


Fig. 4 Main contributors of extracellular signals. (a, top) Postsynaptic currents influence extracellular voltage recordings. Overlay of events elicited by single action potentials of an interneuron and the resulting distribution of unitary postsynaptic amplitudes in rodent hippocampus. (a, bottom) The distributed nature of sinks and sources induced by postsynaptic currents. A single excitatory synapse (*solid circle*) is activated along an apical branch and its impact is propagated along the entirety of extracellular space due to the spatially distributed morphology of the excitatory neuron and the presence of passive return currents. (b, top left) Spike triggered average EAP waveform of a layer 5 pyramidal neuron with the intact EAP waveform (*red*) and when the EAP negativity is missing (*black*); window of 0.6 ms around spike initiation time is substituted by a spline. (Top right) Extracellular traces composed using the L5 pyramidal EAP waveform (*red*); using the intact EAP waveform; *black* using the de-spiked EAP waveform). As observed, the typical EAP-related negativity is missing whereas the remainder of the waveform is attributed to slow afterpotential currents. (Bottom) Mean spectral density as a function of temporal frequency for the intact and de-spiked EAP waveform of the L5 pyramidal (*green*: no spiking; *red*: 1 Hz; *black*: 8 Hz; *blue*: 30 Hz). As observed, the effect of spike afterpotentials (even in the absence of the salient EAP-negativity) can impact voltage recordings in frequencies as low as ~20 Hz. (c) Impact of Ca-dependent dendritic spikes on extracellular voltage recordings. A computational model of a layer 5 pyramidal neuron in the presence (*left*) and absence (*right*) of the Ca-hot-zone (location shown by the *arrow*) is used to emulate the electric field produced by a single neuron. The simulated depth LFP for the two cases is shown by the traces. The presence of the Ca-hot-zone and elicitation of a Ca-spike give rise to a strong, long-lasting event in the superficial regions of the cortex. Figure contributions are from (a, top) Bazelot et al. (2010), (a, bottom) Lindén et al. (2010), (b) Anastassiou et al. (2015), (c) simulations by A.S. Shai and C.A. Anastassiou

Beyond the location-related aspects of synaptic input, another important factor crucially dictating the characteristics of extracellular electrophysiology signals is **input correlation**, i.e., the amount of synaptic input impinging along a neuron or neural population in a unit of time. Typically, enhanced input correlation is manifested in larger V_e and local field potential (the lowpass filtered part of V_e or *LFP*) amplitude, even if such intuition is not always warranted (see below). Beyond the extracellular LFP magnitude, an additional feature of electrophysiology recordings impacted by input correlation is the spatial extent or spread. For example, for uncorrelated input the majority of the extracellular voltage signal measured by an electrode originates from neurons within a lateral distance of approximately 200 μm (Katzner et al. 2009; Xing et al. 2009; Lindén et al. 2011; Reimann et al. 2013). Notably, such low input correlation results in the independence of the region size generating the LFP from neural morphology and the spatial distribution of the synapses. In the presence of more considerable input correlation, the picture changes drastically: pyramidal neurons with their extended spatial morphologies as well as their synaptic specialization tend to dominate the extracellular field. Moreover, correlated synaptic inputs give rise to correlated neural membrane sources that result overall in stronger LFP amplitude. Yet, the degree of LFP amplitude enhancement depends on the spatial separation between impinging synaptic currents and return currents—for substantial separation (i.e., spatially inhomogeneous input along the extended dendritic arbor), the LFP amplitude enhancement becomes significant whereas for smaller separation (i.e., spatially homogeneous input along more compact dendritic arbor), LFP enhancement becomes weaker. Such interdependence between neural morphology features, location of synaptic inputs, input correlation, etc., can putatively explain the disparate length scales encoded by extracellular recordings (Kreiman et al. 2006; Liu and Newsome 2006; Katzner et al. 2009; Xing et al. 2009).

Thus far, we have mostly considered chemical communication via dedicated synapses that are prevalent in brain tissue. Another component whose impact on population dynamics and, as a result, the extracellular voltage remains unaccounted for is electrical synapses, which provide a low-resistance pathway between neurons permitting the direct transmission of electrical signals. **Gap junctions** (GJs), the morphological correlate of electrical synapses, have been used as a proxy for electrical coupling and to infer electrically coupled network architectures. Numerous studies have revealed such networks of electrically coupled neurons in many mammalian brain structures. In cortex, extensive coupling has been reported primarily between inhibitory parvalbumin-positive (PV) interneurons and between somatostatin (SST)-expressing neurons. Such PV- and SST-expressing inhibitory neurons critically contribute to many aspects of ensemble encoding in the mammalian brain (Hu et al. 2014), with one of their most prominent roles being balancing excitation and shaping rhythmic activity. In addition, PV interneurons, the most populous among inhibitory cortical cell types, shape cortical ensemble activity, both during gamma (Szabadics et al. 2001; Traub et al. 2001) and during other rhythms and events such as hippocampal theta or sharp waves. Given that proximally located SST and PV neurons are connected both via chemical and

electrical synapses (for a recent review, see Pereda 2014) in the developing and in the developed neocortex (Connors et al. 1983), can GJs alter extracellular electric fields? Because ions passing through GJs do not enter the extracellular space, it follows that GJs themselves contribute neither to the extracellular current flow nor to the extracellular field explicitly. On the other hand, because GJs contribute to the functioning of inhibitory cells and cell populations altering, for example, their spiking characteristics, they can have an implicit effect on field activity that hitherto has remained unexplored.

Active Membrane Currents

Most neurons produce brief action potentials or spikes that travel along their axons and give rise to synaptic currents at the synapses. It is through the propagation of such electric activity from one neuron to its post-synaptic targets that information is generated and processed within neural populations. Action potentials are produced through active ionic membrane mechanisms allowing the exchange of ions such as Na^+ , K^+ and Ca^{2+} across the membrane. Specifically, fast, Na^+ -dependent **spikes and spike afterpotentials** generated at the axon initial segment and somata of neurons give rise to the strongest currents across the neuronal membrane, detected as ‘unit’ or ‘spike’ activity in the extracellular medium. Although Na^+ -spikes generate large-amplitude and transient (typically lasting 0.5–1 ms) V_e deflections proximal to the soma with a cascade of ionic mechanisms, spike- and spike afterpotential-associated fields remain local (Fig. 3c). The fact that spikes typically last less than a few ms has led to the assumption that they only contribute to extracellular unit activity whereas not appreciably to slower signals such as the LFP or the scalp-recorded EEG like the VAN. Yet, synchronously elicited action potentials (e.g., due to increased spike correlation) from many proximal neurons can contribute substantially to slower bands of extracellular recordings (Anastassiou et al. 2015; Belluscio et al. 2012; Schomburg et al. 2012; Taxidis et al. 2015; Zanos et al. 2011). In addition, it has been shown that spikes give rise to slower, smaller-amplitude afterpotential currents. These spike afterpotentials have recently gathered much attention with studies showing that they can impact bands as low as 20 Hz (Fig. 4b; see also sections below).

Another type of active membrane current is constituted by **Ca-spikes and Ca-related signals**. Decades of work, mostly in vitro, have revealed that the dendrites of cortical pyramidal neurons support a variety of nonlinear signals such as so-called NMDA spikes, Ca-spikes, Na-spikelets and backpropagating action potentials. Of particular interest are the temporally extended NMDA spikes and dendritic Ca-spikes. With regards to **NMDA spikes**, basal, oblique, and apical tuft dendrites of cortical pyramidal neurons receive a high density of glutamatergic synaptic contacts. The synchronous activation of 10–50 such neighboring glutamatergic synapses triggers a local dendritic regenerative potential, NMDA spike/plateau, that is characterized by significant local amplitude (40–50 mV) and an extraordinary duration (up to several hundred milliseconds). Notably, the

conductance of the glutamate-dependent NMDA receptor (NMDAR) channel is also dependent on voltage, giving the NMDAR its spiking ability. NMDARs are found on the thin tuft dendrites of pyramidal neurons, such as those that reside in layer 1, and have been shown to support spatially localized ($\sim 30 \mu\text{m}$) all-or-none spiking events, due to the glutamate binding requirement. Given the electrotonic distance to the spike initiation zone, single NMDA-spikes do not, in general, cause somatic action potential output. Yet, the effect of such NMDA spikes depends on the location where they take place: if occurring in the apical tuft, they have the ability to substantially depolarize the entire tuft region whereas, if occurring closer to the cell body, they can depolarize the soma in a fashion similar to an UP-state. In addition, it has been shown that a distributed set of multiple NMDA-spikes across the dendritic tuft has the ability to cause action potentials during in vivo sensory stimulation (Palmer et al. 2014).

Dendritic Ca-spikes are nonlinear events mainly attributed to large conductance, high-voltage activated channels along the pyramidal dendrites that mediate a sustained Ca^{2+} -influx in a variety of dendrites (Larkum et al. 1999, 2009; Shai et al. 2015). The apical dendrite of pyramidal neurons has a main bifurcation that occurs in L2/3 or L1 and contains a high density of voltage-dependent Ca-channels. This “hot spot” of Ca-channels, alongside other nonlinear channels, supports a relatively slow but large all-or-none depolarizing current known as the Ca-spike. Lasting for ~ 20 – 100 ms in in vitro conditions, and possibly longer in vivo, the Ca-spike has the ability to depolarize a pyramidal neuron for an extended period of time. These Ca-spikes can be triggered with a variety of mechanisms: by strong synaptic drive, by a triplet of back-propagating action potentials or via an extracellular stimulus.

Due to the location of the channels responsible for NMDA and Ca-spikes, these signals are well-suited to being controlled and evoked by inputs into the dendritic arbors of excitatory neurons. One such cortical pathway is long-range connections into layer 1. Indeed, channel-rhodopsin-assisted mapping techniques have shown that higher-order areas send strong-direct excitatory input into the apical dendrites of pyramidal neurons (Fig. 5; see also Yang et al. 2013). What are the functional consequences of NMDA and Ca-spikes in the dendrites of a pyramidal neuron? In vitro experiments have shown that the Ca-spike can integrate with a backpropagating action potential to elicit a spike burst (i.e., a multitude of somatic spikes elicited within a few tens of milliseconds) at the soma (Larkum et al. 1999). It is difficult to precisely control the amount and timing of synaptic inputs into spatially segregated areas of a single neuron experimentally, though some efforts deserve recognition (Jarsky et al. 2005). Modeling approaches present themselves as useful tools to explore the possible functional roles of a complicated mixture of linear and nonlinear channels across the dendritic membrane, as well as their interactions with large barrages of synaptic input (e.g., Shai et al. 2014).

After creating a detailed multi-compartmental model of a L5 pyramidal neuron based on a combination of previous modeling work (Hay et al. 2011) and dual soma and dendrite patch clamp recordings in V1, Shai et al. (2015) imposed barrages of dendritic and somatic excitatory synapses onto a single cell. The results of this simulation showed that the coincident input of perisomatic and apical input elicited

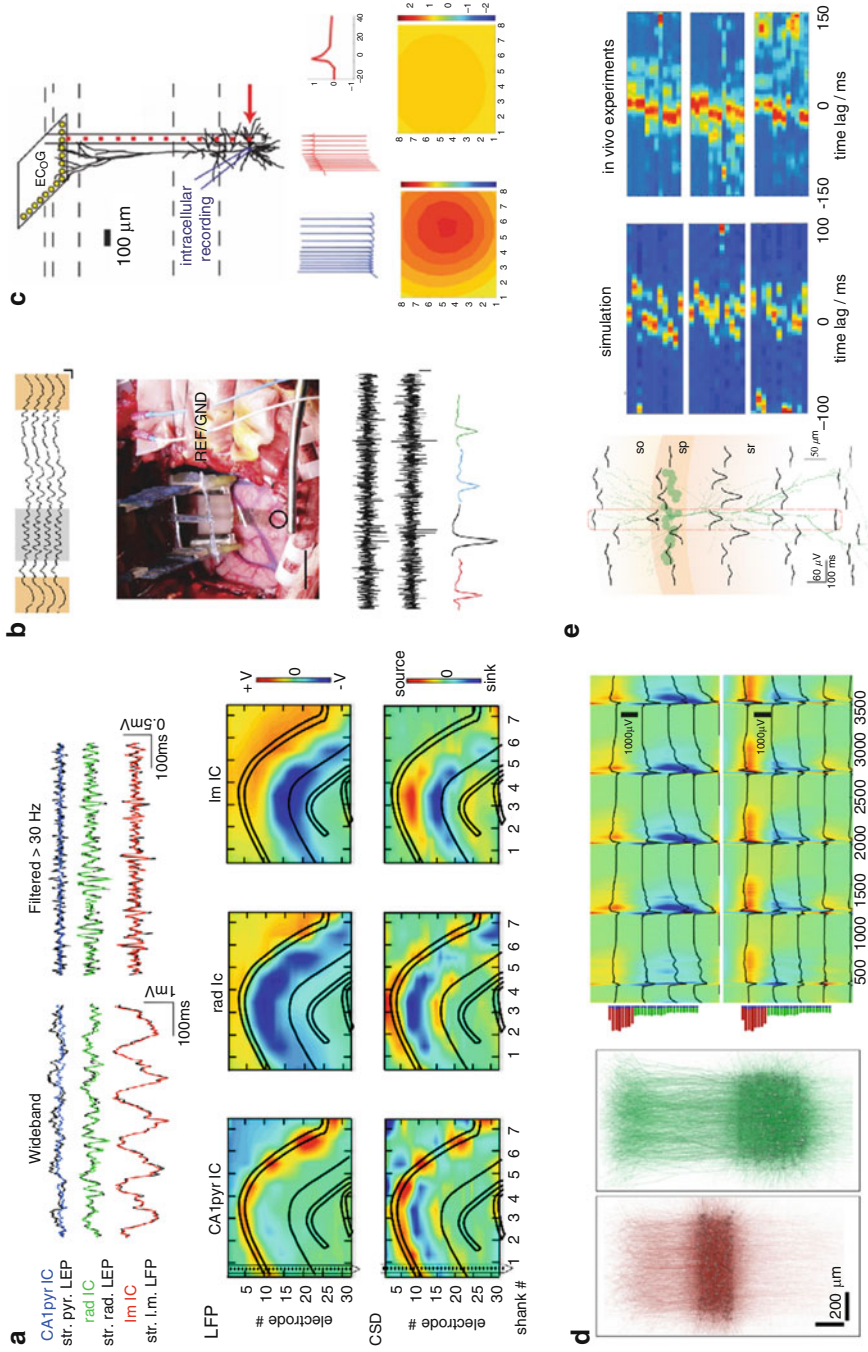


Fig. 5 Monitoring the electrical activity of the brain. (a, top) The traces of wide-band recordings (left) and 30 Hz highpass filtered (right) hippocampal CA1 shown together with their independent components (obtained with independent component analysis). (Bottom) Two-dimensional voltage (LFP) and CSD

a burst of high-frequency action potentials at the soma, whereas only perisomatic or apical input in isolation would not. Furthermore, this effect was dependent on the Ca hot-spot around the bifurcation point. This finding was further summarized in a simple abstract model whereby input into dendrites of the abstracted cell would modulate, in a thresholded manner, the input-output relationship between perisomatic input and output frequency. This “composite-sigmoid” model captured the complicated interaction of perisomatically elicited backpropagating action potentials (mediated by nonlinear sodium channels in the apical trunk) and the tuft elicited Ca-spike, in a compact form, thus elucidating the computational structure of a single pyramidal neuron.

In particular, the Ca-spike, being a relatively slow and large-amplitude signal in the superficial layers, may constitute a candidate for NCC. Multi-compartmental modeling is well suited to disentangling the potential contributions of different cell types of channels to extracellular signals (e.g., Reimann et al. 2013; Fig. 4c). Although more thorough investigations are needed, both in terms of modeling and experiments, to establish the role of the Ca-spike in the superficial extracellular signature, these simulations show that the Ca-spike is indeed a plausible mechanism for these signals (Fig. 6).

Fig. 5 (continued) maps of the three main CA1 independent components from a rat with an electrode array spanning the transverse axis of the hippocampus (seven shanks spaced 300 mm; one shank shown on the *left*) indicates activation of different projections (*CA1pyr* pyramidal layer, *rad* radiatum, *lm* lacunosum moleculare). (b) Electrocorticography (ECoG) records indicating periods of behavior-relevant slow oscillations (*orange*) and spindles (*gray*). (Bottom) Intraoperative ECoG recordings in human patients using new technologies have the ability to detect spiking. Highpass filtered traces from a novel 64-grid electrode containing spiking activity (*black traces*). Below, sample spike waveforms are shown. (c) Simulation of an individual neuron (layer 5 pyramidal injected with intracellular somatic current by a pipette: intracellular somatic spiking shown in *blue* is detected in the extracellular space by a proximal electrode (*red*; part of a silicon depth electrode) as well as by the ECoG strip electrode (simulating the same layout as the one in panel b)). The spike-triggered average ECoG signal from the middle of the ECoG strip is shown (*right*). (Bottom) The spike triggered average ECoG field for two cell types extending to superficial layers: a layer 23 pyramidal (*left*) and a layer 5 pyramidal neuron (*right*). While the amplitude of the spiking ECoG signature is very similar, the spatial extent is markedly different. (d, *left*) A large-scale, biophysically realistic model of thousands of reconstructed and interconnected neocortical layer 4 (*red*) and layer 5 (*green*) pyramidal neurons emulating a patch of deep cortical volume. The population model was used to study the extent to which active membrane conductances impact the extracellular LFP and CSD signals. (*Right*) Two scenarios were instantiated: passive-only membranes and active ones. The simulated LFPs and CSDs show the result of these simulations (*top*: passive-only; *bottom*: active) with the spatiotemporal characteristics of the LFP and CSD being markedly different. (e, *left*) Hippocampal model of the CA1 region consisting of reconstructed excitatory neurons capturing the various projections during sharp wave ripples accounts for the extracellular signals during such events. (*Right*) Replay sequences during sharp waves yield consistent LFP patterns in the ripple (150–200 Hz) bandwidth. As observed, simulations point to the spatiotemporal patterned activity that is also observed in the same band in vivo, reflecting the spiking activity of cell assemblies activated during sharp waves. Figure contributions are from (a) Schomburg et al. (2014), (b) Khodagholy et al. (2015), (c) simulations by C.A. Anastassiou and A.S. Shai, (d) Reimann et al. (2013), (e) Taxisidis et al. (2015)

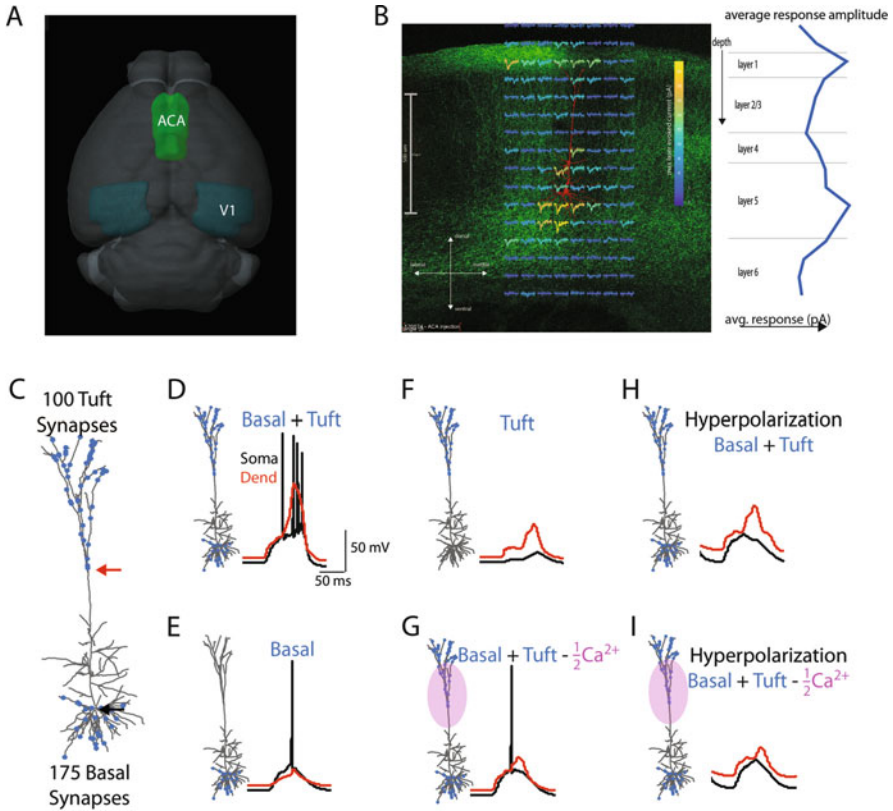


Fig. 6 A mechanism of coincidence detection via feedback into layer 1. (a) A *top view* of the mouse brain showing the anterior cingulate cortex (ACA, a frontal region) and primary visual cortex (V1). (b) The anterograde projections of ACA axons into V1 show a clear striation in layer 1 (green fluorescence). Subcellular channel-rhodopsin-assisted mapping (sCRACM) on a layer 5 pyramidal neuron (*red*) shows strong excitatory input into the apical tuft dendrites. (c) 100 tuft and 175 basal NMDA/AMPA synapses are distributed randomly across the apical tuft and basal dendrites of a multi-compartmental L5 pyramidal neuron model. All synapses are randomly and uniformly elicited in time across 100 ms. In the following *panels*, somatic traces are in *black* and dendritic (location shown by the *red arrow* in *c*), are in *red*. (d) Simultaneous tuft and basal inputs trigger a burst of somatic action potentials and a dendritic Ca^{2+} spike, whereas (e) basal inputs alone evoke only a single somatic spike. (f) Apical tuft inputs alone do not evoke somatic spiking. (g) Reducing Ca^{2+} channel conductance by 50 % during tuft and basal input gives rise to a single somatic spike. (h) When applying a 200 pA hyperpolarizing DC current to the soma, the subthreshold response of the tuft and basal inputs are similar to the case with Ca^{2+} -conductances reduced shown in (i), even though the suprathreshold (b, c) cases are remarkably different. (a) Taken from the Allen Institute Brain Explorer. (b) Experiments performed by Adam Shai, but also see Yang et al. (2013), for similar results. (c–i) Taken with permission from Shai et al. (2015)

Non-neural Contributors (Glia, etc.)

During the last two decades, glial cells have been shown to be of great significance for brain signaling (Volterra and Meldolesi 2005) while also possessing active ionic conductances that result in fairly slow but prominent transmembrane processes being activated during neural activity (Perea and Araque 2007). Electrically passive astrocytes coexist with others that show voltage-dependent currents such as inwardly rectifying or delayed, outwardly rectifying K^+ or both types (D'Ambrosio et al. 1998). Given the abundance of glia in brain tissue, how do these contribute to the extracellular electric field (Wang et al. 2006)? Can certain LFP or EEG bands (such as the slow 0.1–1 Hz band) be influenced by glial and astrocytic transmembrane activity? Such questions are also related to the link between LFP activity, the blood oxygen-level dependent (BOLD) signal and the overall metabolic demands of specific brain areas. Interestingly, the BOLD signal, which has been linked to neural as well as astrocytic activity, has been found to correlate preferentially with specific LFP bandwidths.

Monitoring Neural Activity

In this section of the chapter we present the most prominent methods of monitoring brain activity. We separate this section into two parts: a part on monitoring *spatially local* brain activity and a part on methods used to monitor *spatially extended* (even whole-brain) activity. While local monitoring can offer superior spatiotemporal resolution from identified signal sources, spatially diffuse monitoring offers insights from multiple brain regions, as discussed previously, such distributed processing has been often implied to be a cornerstone for the formation of conscious percepts.

Local Monitoring

Local monitoring refers to measurements of neural activity within fairly small volumes of brain tissue. Historically, the most prominent of these local monitoring techniques have been extracellular voltage recordings either from a single location (e.g., via a metal wire) or multiple locations (multiple wires bundled together or multiple contacts manufactured along the length of a silicon shank; Buzsáki 2004). The exact sampling volume of such extracellular electrodes remains the object of investigations. In the past, various distances have been suggested ranging from a few tens to a few hundreds of micrometers. Here we present and succinctly discuss the most prominent methods (also beyond extracellular recordings) that allow monitoring neural activity within similar volumes.

Action potential elicitation of neurons proximal to a recording electrode is typically reflected in the approximately 1 ms long and 50–200 μV deep negativities of the extracellular voltage time series. These rapid and spatially localized V_e -deflections reflect membrane currents in the axonal and perisomatic region (Fig. 3b, c) but are also impacted by more distributed currents such as backpropagating action potentials traveling along the dendritic arbor of a neuron. An important aspect of monitoring spiking activity in the extracellular space is the inherent, activity-dependent variability of the EAP waveform. Indeed, the EAP amplitude of hippocampal pyramidal neurons can vary as much as 60 % during a high-frequency (approximately 200 Hz) burst (Buzsáki et al. 1996) for example, within a place field (Harris et al. 2001). These features, together with artefactual sources of variability from electromyogram contamination or hardware sources, pose challenges for spike waveform-based clustering and classification of neurons in vivo. Specifically, the most salient features of EAPs, the EAP negativity attributed to fast sodium- and potassium-dependent currents and the immediately proceeding EAP positivity attributed to slower potassium (but also calcium) currents, can substantially vary as a function of spike frequency. In general, the spike variability of the EAP amplitude is more pronounced compared with intracellular spike variability (Anastassiou et al. 2015; Henze et al. 2000) and is non-monotonic as a function of spike frequency. The temporal EAP waveform features, such as halfwidth or decay time of spike repolarization, vary more reliably with the intracellular waveform than the amplitude (Anastassiou et al. 2015; Barthó et al. 2004). It follows that, while EAP recordings are fairly straightforward to obtain compared to more elaborate methodologies, can be performed in deeper brain structures and offer superior temporal resolution, their disadvantage lies in the difficulty of the procedure involved in separating spikes and EAP waveforms originating from different neurons, a process often referred to as spike clustering.

When the wideband signal from extracellular recordings is lowpass filtered (typically below 300 Hz), the resulting time series is referred to as the *local field potential* (LFP). The LFP has been studied extensively in the past as, in contrary to highpass part of the recordings, it reflects electrogenesis from a spatially more distributed region (Fig. 5a). What the length scale of the LFP is has been a debated topic and remains a vibrant field of science (e.g., Katzner et al. 2009; Xing et al. 2009; Ray and Maunsell 2011; Lindén et al. 2011; Reimann et al. 2013). What is clear is that the spatial extent of the LFP is not a static feature depending on multiple factors such as cytoarchitecture, input correlations (see above), etc. Hitherto, the LFP has been considered to mainly reflect postsynaptic and associated return currents (even if recently more LFP-contributors have been identified); as such, it is uniquely positioned to measure input into and output from a particular brain region (e.g., Mitzdorf 1985; Colgin et al. 2009; Einevoll et al. 2013; Logothetis et al. 2007; Buzsáki et al. 2012; Schomberg et al. 2014; Taxidis et al. 2015).

A method used in conjunction with LFP recordings is the so-called *current source density* (CSD) analysis. The CSD analysis is a particularly useful tool in deciphering the location of the extracellular current sources and sinks giving rise to

the LFP (Fig. 5a). CSD per se represents the volume density of the net current entering or leaving the extracellular space (Nicholson and Freeman 1975; Mitzdorf 1985; Buzsáki et al. 2012). Unfortunately, it is not possible to conclude from the CSD analysis alone whether, for example, an outward current close to the cell body layer is due to active inhibitory synaptic currents or reflects the passive return current of active excitatory currents impinging along the dendritic arbor. Such insights have to be gathered from complementary information such as the cytoarchitecture of the brain region under investigation, its anatomy, projecting pathways, etc. Even so, CSD analysis can point to regions of interest to be studied more elaborately.

Conventionally it has been thought that spiking currents cannot affect temporally slower signals such as the LFP or the EEG due to the rapid, approximately 1-ms transient sodium/potassium charge transfer giving rise to the stereotypical intracellular positivity (or extracellular negativity). Lately this view has been challenged by a number of studies showing that neural spiking can affect electric signals at much lower frequencies than the typical time scales suggested by action potentials (Belluscio et al. 2012; Zanos et al. 2011; Ray and Maunsell 2011; Schomburg et al. 2012; Reimann et al. 2013; Anastassiou et al. 2015). What part of the EAP waveform can impact power at slow bands of extracellular recordings? This has been the focus of a few studies (e.g., Zanos et al. 2011; Belluscio et al. 2012; Anastassiou et al. 2015). In a recent one, the authors performed so-called “de-spiking,” i.e., the procedure of substituting a window of 0.6 ms before and after spike initiation time with a different (non-spiking) time series in the extracellular voltage time series (Belluscio et al. 2012), in experiments where both the intracellular and extracellular spikes were monitored concurrently (Anastassiou et al. 2015). This resulted in EAP waveforms lacking the typical spike negativity but containing the characteristic afterpotential repolarization. Performing spectral analyses of the de-spiked time series led to a surprising conclusion: spike afterpotential currents of pyramidal neurons can impacted the spectrum of recorded signals as low as 20 Hz, i.e., bands hitherto solely related to synaptic processing (Fig. 4b). Importantly, when the same analyses using the EAP waveform from basket cells was performed, the outcome was very different: spiking of these neurons minimally contributed to spectral power under 100 Hz and, even then, did so only for elevated spike frequencies. The lack of impact of basket cell spiking to LFPs under 100 Hz was attributed to their temporally narrow EAP waveform as well as the lack of long-lasting depolarizing currents (compared to pyramidal neurons). The study concluded that the effect of EAPs in such low frequencies was attributed to the slower, smaller amplitude repolarization typically attributed to slower potassium- and calcium-dependent currents difficult to distinguish in vivo.

Electrocorticography (ECoG) is the intracranial recording of electrophysiological signals using electrodes and multi-electrode arrangements (grids) from the surface of the brain after craniotomy and has been used for decades to monitor (and sometimes perturb) cortical activity. Specifically, ECoG recordings have conventionally been used to record slow signals (similar to the LFP) related to

brain states or evoked activity, though spiking activity has been difficult to detect (Fig. 5b). In that sense, ECoG has been largely used as a spatially distributed monitoring method much related to electroencephalography and magnetoencephalography (see below). Yet, very recently, advances in technology and materials have for the first time allowed robust recording of cortical spiking (Khodagholy et al. 2015) using ECoG (Fig. 5b, c), rendering the possibility of concurrent monitoring of intra- and inter-cortical processing in terms of spiking and slower activity from the brain surface.

Beyond electric recording methodologies, optical imaging techniques capturing electric or ionic activity in neurons have flourished over the past decade or so. Specifically, voltage changes can also be detected by membrane-bound voltage-sensitive dyes or by genetically expressed voltage-sensitive proteins (Siegel and Isacoff 1997; Grinvald and Hildesheim 2004; Akemann et al. 2010). Using the *voltage-sensitive dye imaging* (VSDI) method, the membrane voltage changes of neurons in a region of interest can be detected optically, using a high-resolution fast-speed digital camera, at the excitation wavelength of the dye. A major advantage of VSDI is that it directly measures localized transmembrane voltage changes, as opposed to the extracellular potential. A second advantage is that the provenance of the signal can be identified if a known promoter is used to express the voltage-sensitive protein. Limitations are inherent in all optical probe-based methods (Denk et al. 1994); for VSDI these include interference with the physiological functions of the cell membrane, phototoxicity, a low signal-to-noise ratio and the fact that it can only measure surface events.

Calcium imaging has emerged as a promising technology for observing hundreds to thousands of neurons within a micro-circuit with both high spatial resolution and precise localization to specific brain regions. The technique works by introducing calcium-sensitive indicators into neural populations of interest and then imaging these neurons in vivo through a light microscope. These fluorescence measurements are interpreted as a proxy for the underlying neural spiking activity, as there is a biological relationship between elicited action potentials and changes in calcium concentration; a spike causes increases in $[Ca^{2+}]$, which gradually decays due to cell buffering and other extrusion mechanisms. A major advantage of Ca-imaging is that, in combination with genetically modified cre-animals, it offers the ability to record activity from different cell types. In addition, fluorophore kinetics have been drastically reduced so that, in principle, single-spike resolution is obtainable in a limited volume. On the other hand, a major problem arises when intending to monitor spiking activity in larger volumes; instead, what is recorded is a noisy and temporally sub-sampled version of the spiking activity, which in some cases can be orders of magnitude slower than the underlying neural dynamics. Even so, technology advances are continuously offering indicators with faster response times and increased signal-to-noise ratio.

Finally, a method recently revamped as a test bed for understanding the origin and functionality of signals is *computational modeling*. The first model to link intra- and extracellular voltages was the work of Pitts (1952) describing the extracellular negativity appearing as a result of spiking. Accordingly, the first simulations shedding light into the LFP signal were the pioneering work by Shepherd and Rall explaining the LFP recordings in the olfactory bulb of rabbit from first principles

(Rall and Shepherd 1968). Since that time, a number of significant contributions have been made with respect to the neural underpinning of brain signals, where more involved computational models have been employed, for example, accounting for different cell types, varying ratio of excitation and inhibition, etc.

A caveat of simulations typically used to study brain functioning and recreate brain signals is that they have remained somewhat too conceptual. Neurons are typically taken as point-like processes with rules of connectivity imposed upon such nodes. While such simulations have proven informative with regards to analyzing network dynamics (Koch 2004), signals related to electric field activity are induced by the multitude of synaptic and membrane conductances activated along the intricate, three-dimensional morphology of neurons (see also previous sections) and are critically impacted by factors such as the alignment of dendrites and other neural processes, input impinging along these processes, etc. (see above; Buzsáki et al. 2012). Thus, the use of point neurons, while informative for illuminating computational principles, either presumes or even fully neglects the primary means by which such effects are mediated, that is, ionic fluxes along the neural membrane and the extracellular medium. These restrictions are by no means limited to models of electric activity (Fig. 5d, e). For example, a similar lack of understanding is combined with models attempting to replicate Ca-imaging response. In this case, limitations do not arise from the lack of morphology features anymore but instead from the lack of understanding and accurate representation between intracellular Ca-dynamics and the resulting fluorescence signal.

The recent rise in computational power and advances in parallelization have allowed larger, more realistic models to be implemented. Such models carry the potential of being able to link subcellular and cellular biophysics with locally measured signals such as cortical spiking, LFPs, Ca-imaging, etc. For example, morphologically detailed and biophysically realistic single-neuron (Gold et al. 2006; Druckmann et al. 2007; Hay et al. 2011) and population models (Pettersen and Einevoll 2008; Lindén et al. 2011; Schomburg et al. 2012; Reimann et al. 2013; Taxis et al. 2015) have offered considerable insights into extracellular spiking and LFP signals. Even more recently, large-scale simulation programs combining unprecedented level of detail have been initialized promising to unravel novel insights into a plethora of brain signals (e.g., Markram et al. 2015).

Spatially Distributed Monitoring

Spatially distributed monitoring relies on the same biophysical principles as local monitoring, yet the inability of measuring highly localized sinks and sources due to spatially undersampling renders the origins of electric activity spatially equally diffuse.

Electroencephalography (EEG) is one of the oldest and most widely used methods for the investigation of the electric activity of the brain (Niedermeyer and Lopes da Silva 2005; Nunez and Srinivasan 2006). The scalp electroencephalogram is a spatiotemporally smoothed version of the ECoG (though the impact of the skull on the recorded signal needs also to be accounted for) or LFP, integrated

over an area of 10 cm² or more. Under most conditions, it has little discernible relationship with the firing patterns of the contributing individual neurons, largely due to the distorting and attenuating effects of the soft and hard tissues between the current source and the recording electrode. The recently introduced ‘high-density’ EEG recordings, in combination with source modelling that can account for the gyri and sulci (as inferred from structural MRI imaging) of the subject, have substantially improved the spatial resolution of EEG (Nunez and Srinivasan 2006; Ebersole and Ebersole 2010).

Magnetoencephalography (MEG) uses superconducting quantum interference devices (SQUIDs) to measure tiny magnetic fields outside the skull (typically in the 10–1000 fT range) from currents generated by the neurons (Hämäläinen et al. 1993). Because MEG is non-invasive and has a relatively high spatiotemporal resolution (~1 ms, and 2–3 mm in principle), it has become a popular method for monitoring neuronal activity in the human brain. An advantage of MEG is that magnetic signals are much less dependent on the conductivity of the extracellular space than EEG. The scaling properties (that is, the frequency versus power relationship) of EEG and MEG often show differences, typically in the higher-frequency bands, that have been attributed to capacitive properties of the extracellular medium (such as skin and scalp muscles) that distort the EEG signal but not the MEG signal (Dehghani et al. 2010).

Functional magnetic resonance imaging (fMRI) is an imaging technique that monitors oxygenation levels of blood flow in the brains of animals and humans. Specifically, the BOLD contrast has been used as a proxy for neural activity, though the exact relationship between neural processing and the output signal is a complicated one (Logothetis and Wandell 2004). A number of pivotal studies have appeared over the years relating the BOLD signal with depth LFP recordings rather than spiking (Logothetis et al. 2001; Logothetis and Wandell 2004; Nir et al. 2007; Schölvinck et al. 2010). The main advantage of fMRI is that it can be applied in a brain-wide fashion, allowing for whole-brain associations, and it is non-invasive. At the same time, the temporal sampling rate is fairly slow (typically fractions or a few Hz) and the voxel size of the signal acquisition is considerable (e.g., from fractions to a few mm).

Linking spatially distributed measurements with the biophysics and workings of networks and circuits all the way to single-cell and synaptic contributions typically measured via local measurements has remained a challenge, mainly due to the multiple spatiotemporal scales involved requiring simultaneous monitoring at all levels. While such monitoring is difficult to pursue in humans, recent advances in sensing technology have allowed performing it in other animals, particularly rodents. For example, as mentioned earlier, recent advances in material and technology have allowed simultaneous measurement of spiking, LFPs and ECoG in rodents (but also humans), offering the possibility to link between micro-, meso- and macroscopic electric signals (Khodagholy et al. 2015). In similar fashion, the relationship between the BOLD fMRI signal has been studied in conjunction with spiking and LFP measurements (e.g., Logothetis et al. 2001; Nir et al. 2007; Whittingstall and Logothetis 2009) and, recently, by engaging specific neural population via optical perturbation (Lee et al. 2010).

Computational modeling has the ability to link across scales and relate microscopic with meso- and macroscopic observables. Yet, at the level of distributed brain circuits, detailed representations of each circuit and its elements—such as synapses or single-neuron morphologies—becomes prohibitive. Even so, more abstract models of neural processing, such as circuits consisting of leaky-integrate-and-fire units, have provided many insights into the functioning of distributed brain circuits during sleep and wakefulness (Hill and Tononi 2005), the perception-action cycle (Eliasmith et al. 2012), etc. With regards to conscious perception, modeling has been employed in attempts to link between the various signals and neural dynamics during tasks. In an important study, Dehaene and colleagues (2003) used a neural network model to investigate mechanisms underlying visual perception typically giving rise to activity patterns such as sustained activity in V1, amplification of perceptual processing, correlation across distant regions, joint parietal, frontal, and cingulate activation, band oscillations, and the p3b waveform. The neural network model indicated that access awareness (the step of conscious perception) is related to the entry of processed visual stimuli into a global brain state that links distant areas, including the prefrontal cortex, through reciprocal connections and thus makes perceptual information reportable by multiple means. This study is an excellent example of the kinds of insights computational modeling can offer towards relating signals linked to conscious processing with underlying neural processing in distributed areas.

From Cellular Biophysics to Systems and Computations Associated with Conscious Perception

In the previous section, we reviewed a single-cell mechanism for spike bursting via the dendritic Ca-spike of pyramidal neurons, whose extracellular signature is a plausible candidate for a late superficial current sink. Cortical layer 1 is unique in that it is extremely sparse, and the vast majority (upwards of 90 %; Hubel 1982) of the synapses there are from long-range inputs rather than from the local circuit. Importantly, the pyramidal neurons whose dendrites support Ca-spikes are precisely those neurons that make long-range connections themselves, both cortically (feedforward, horizontal, and feedback⁷) and subcortically. What computational role could be played by such a physiological and anatomical setup?

One intriguing possibility, which we will call Association by Apical Amplification (AAA), was described by Matthew Larkum (2013). AAA takes a largely bottom-up approach, starting from the detailed physiology of pyramidal neurons and the anatomy of long-range connections in the cortex. Of particular importance is the laminar structure of long-range feedforward and feedback axons in the cortex. There is now ample evidence that feedforward connections strongly innervate the

⁷There is an “indirect” pathway for cortico-cortico information flow through the thalamus, and some argue that this might be the main way that information is transferred from one area of cortex to another (Sherman and Guillery 2011).

basal dendrites of layer 5 pyramidal neurons with excitatory synapses. Feedback axons innervate layer 1, where the dendritic tufts of pyramidal neurons reside. As discussed previously, the physiology of layer 5 pyramidal neurons allows for a coincidence detection mechanism, whereby concurrent excitatory input into both the basal and apical tuft dendrites causes a high frequency burst.

Additionally, the local inhibitory circuit consists of a number of different cell types that can generally be classified into distinct groups based on their specific effects on either the somatic or apical areas of the pyramidal neuron. For instance, neurogliaform cells in layer 1 metabotroically inhibit voltage-gated calcium channels in the apical dendrites (Palmer et al. 2012; Pérez-Garci et al. 2013), whereas single bouquet cells in layer 1 disinhibit the apical dendrite via their inhibitory effects on layer 2/3 inhibitory cells (Jiang et al. 2013). SST-positive inhibitory neurons are known to directly inhibit the apical dendrites, whereas PV-positive inhibitory neurons directly inhibit the basal dendrites, affording these groups of neurons distinct computational roles in the regulation of pyramidal neuron output (Royer et al. 2012; Shai et al. 2014). In this way, inhibition of the apical dendrites by neurogliaform cells or SST-positive interneurons can act as a form of gain control, regulating the frequency of firing in pyramidal neurons. Alternatively, inhibition of neurogliaform cells, for instance via cholinergic action (though under certain conditions, acetylcholine can have opposite effects; Brombas et al. 2014), can bias pyramidal neurons to high frequency firing.

Taken together, pyramidal neurons and the local inhibitory circuit that surrounds them are well suited to associate feedback and feedforward information streams (Fig. 7). That association, signaled via a high-frequency spike burst in a pyramidal neuron, is then communicated to other areas of the brain, including other areas of cortex. For instance, a pyramidal neuron receiving feedforward orientation information from V1 and motion information via feedback from V5 can bind these two information streams. These associated signals can then contribute, via their influence on the apical or basal dendrites of far-away pyramidal neurons, to other associations. The single-cell mechanism through which concurrent basal and tuft excitatory input creates spike bursting has been named the BAC mechanism (Larkum 2013). In this way, the BAC mechanism causes high-frequency burst firing, as is often observed in vivo (de Kock and Sakmann 2008; Buzsáki and Mizuseki 2014), whereas input into only the basal dendrites will only cause tonic low-frequency firing (Fig. 5). Long-range input can also robustly regulate the BAC mechanism indirectly by recruiting the effect of the different cell types in local inhibitory circuit.

Importantly, low-frequency firing is still available as a unit of information transfer in cases where excitatory input exists into the basal dendrites in the absence of excitation in the apical tufts or when the BAC mechanism is inhibited. These different modes of firing (low-frequency vs. high frequency bursting) can have substantially different influences postsynaptically (Buzsáki and Mizuseki 2014; Lisman 1997). For instance, different short-term plasticity mechanisms act as filters, allowing only certain frequencies to effectively communicate with downstream neurons (Markram et al. 1998; Tsodyks and Markram 1997; Tsodyks et al. 1998). There is evidence that presynaptic bursts cause postsynaptic potentials with substantially greater efficacy (>90 %) than single action potentials (~40 %;

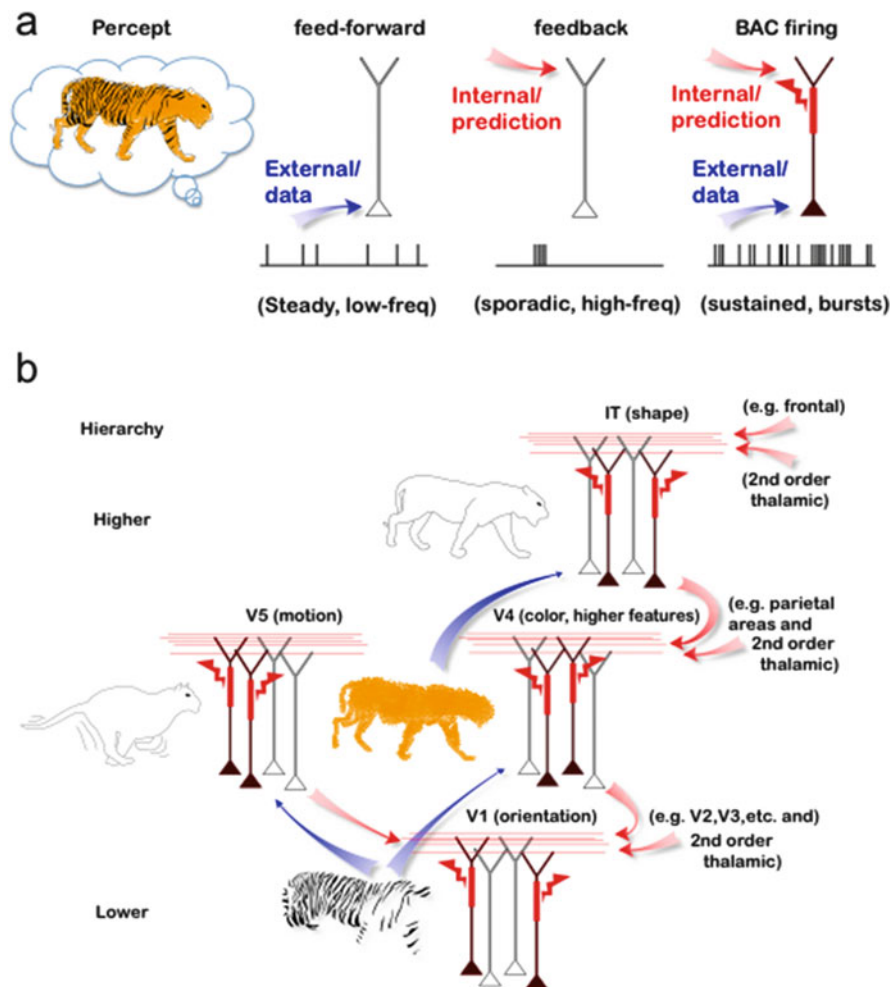


Fig. 7 Association by apical amplification (AAA) connects physiological and anatomical details to network level computation and perceptual representation in the cortex. (a) As shown in Fig. 6, input into the basal dendrites of a cell causes steady low-frequency firing in a pyramidal neuron. This feedforward input into the basal dendrites, when combined with feedback input into the apical tufts, causes high frequency burst firing. In the scheme of AAA, feedforward input into the basal dendrites carries sensory information from the periphery, while feedback input into the apical tufts carries predictive information about the stimulus. (b) The parallel feedforward/feedback interactions in multiple areas acts as a selection mechanism to choose which pyramidal neurons are in a state of high frequency firing, ultimately binding different aspects to represent the percept, in this case a tiger (figure from Larkum 2013)

Lisman 1997). In this way, the coincident excitatory input into a pyramidal neuron, representing the association of information from different areas of cortex, can create a unique signal that has markedly different influence on other cortical areas than the integration of a purely feedforward (basal dendrite) input.

AAA importantly serves as a concrete mechanism linking details of electrophysiology and anatomy to larger-scale concepts like perceptual binding and the representation of conscious percepts. The connection between forming a bound representation and consciousness are matters of current debate and require some discussion. The topic of binding is generally separated into two issues (Revonsuo 1999). The first concerns the grouping of sensory inputs to form distinct objects (e.g., combining color, shape, and motion to represent a tiger); the second concerns the inherent grouping of the phenomenal mind to give a single conscious experience (a property called *the unity of consciousness*). The difference between these two sets of concerns is also contentious. One way to think of the difference between these two issues is that the first relates to computational function whereas the second relates to phenomenology. Another interpretation is that the first relates to defining distinct objects and the second refers to arranging those objects into some kind of unified structure. This interpretation holds great relevance to the ideas discussed in the final section. Yet another interpretation is that there is no real difference between the two sets of concerns or even that the unity of consciousness is illusory (Dennett and Kinsbourne 1992). For the purposes of this chapter, it is important to point out that for AAA to be a mechanism of consciousness, it assumes that the functional substrate of phenomenology is the binding of sensory information to form unified wholes.

Let us take into account the different facts that are put together here. The most basic starts with the distribution of nonlinear channels in the apical dendrites of pyramidal neurons, which support nonlinear regenerative spiking, and acts as a mechanism for high frequency burst firing in those cells. Anatomically, these cells make long range connections, both in feedforward pathways where they synapse onto the basal dendrites of other pyramidal neurons, and in feedback pathways, where they synapse into layer 1, and they can act to manipulate the apical dendrites of pyramidal neurons. The extracellular signature of such manipulation, in particular, the dendritic Ca spike, is a large current sink in the upper layers. Psychophysics experiments have found that such a signal correlates to conscious perception. In terms of cortical computation, the association of feedforward and feedback signals might act to bind different aspects of a percept together, though the exact details of such a process at the network level remain elusive. In the next section we discuss candidate theoretical frameworks that might be able to describe such a process in cortex.

Towards a (Unifying) Theory/Framework of Conscious Perception

Before delving into the details of theoretical considerations, it will be useful to quickly review what has been covered in this chapter so far. We began by looking at psychophysical results describing signals that correlate to conscious perception. In particular, late extracellular voltage signals that occur in the superficial layers as well as distributed information processing between different cortical regions were

presented as candidate NCC. From there we considered the physical origins of these extracellular signals, residing in the transmembrane currents brought about by the electrical structure of dendrites and synapses. Dendrites of pyramidal neurons, supporting highly nonlinear NMDA and Ca-spiking, were presented as a likely origin for late extracellular signals in the superficial layers. Next, we asked what computational role such an electrogenic structure could play in terms of single neuron processing of synaptic inputs, and we discussed how pyramidal neurons and their dendrites act as coincidence detectors between inputs into the basal and apical dendrites and additionally have powerful mechanisms to regulate such a coincidence mechanism. Importantly, the output of this single cell mechanism is given by a nonlinear increase in the frequency of action potential outputs, in the form of a burst at 100 Hz or greater. As discussed elsewhere (Larkum 2013) the network implication of such a single cell mechanism is a general principle by which pyramidal neurons distributed across the cortex can be associated with each other, ultimately serving as the physical representation of any given conscious percept.

This series of connections—from psychology to signals, signals to neural biophysics, from biophysics to single cell computation, and single cell computation to network level computation—is built upon more than a century of work in a variety of fields. Still, the connections between these levels of understanding require substantial amounts of work to be sufficiently fleshed out before becoming widely agreed-upon scientific fact. Instead, what has been presented so far should be understood as an attempt to combine results from psychology to physiology in a coherent and testable framework. The testability of this framework is of special import, as this requires (in the best case) taking the somewhat ineffable topic of consciousness into the realm of neurons and their functions.

As an important part of that project, a number of theoretical (and often mathematical) frameworks emerged attempting to describe the abstract underpinnings of representation and consciousness in the brain, ultimately providing a description for what it means, in terms of algorithm or function, to create a representation or to be conscious. In the subsection that follows, we will discuss some of these frameworks and explore how they might be related to the ideas mentioned so far. This discussion will not be an in-depth review but will instead feature a largely conceptual overview. Importantly, the discussion that follows should not be interpreted as arguing for an equivalence between these various theories. Instead, what follows is a discussion of the potential areas of conceptual overlap between seemingly disparate ideas and how they might be brought together, at least at certain points of conceptual intersection.

We will frame this section with Friedrich Hayek's contributions to theoretical psychology, most explicitly given in his 1953 work *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology*. The reasons for this are multifold. First, Hayek's contributions mark a stark departure from multiple theoretical frameworks of that time, for instance behaviorism⁸ and the theory of psycho-

⁸In its' most extreme form behaviorism studies the link between sensory input and behavioral output, and denies that anything is really going on in the mind.

physical parallelism,⁹ ultimately arriving at the modern understanding of the role of the brain in perception. Second, as we will see, there are direct conceptual parallels between his ideas and many of the more mathematically rigorous modern ideas. Third, Hayek's work in theoretical psychology is underappreciated, especially given both its breadth and depth. We will see that Hayek's work provides a conceptual framework that suggests overlap between a number of modern theoretical ideas and AAA (Fig. 8d). With regards to AAA, the main point here is the connection between computation at the single cell level (e.g., as discussed, coincidence-detection, association) and more network-level implications. This link is what Hayek explores.

Hayek's foundational idea is quite straightforward. He posited three orders: (1) the external world (which he called the physical order), (2) the brain (which he called the sensory order), and (3) the mind (which he called the phenomenal order), and he focused his efforts on understanding the relationship between the three. In Hayek's formulation, the state of the brain has an isomorphic correspondence with that of the mind. The structure of the psychological realm, for Hayek, was relational (e.g., psychological objects are defined relative to other psychological objects), and as such, that structure of relationships that make up the psyche had to be recapitulated in the structure of the neural network and its activity. This strict correspondence contrasts with the correspondence between the outside world and the structure of the brain (and thus the mind), which is imperfect, as shown by the existence of sensory illusions. The problem for Hayek was then to describe how the relational network that is the psyche can be encoded in the structure and activity of a neural network, given the computational properties of single neurons that make up that network. Although this might seem trivial to today's standards, it cannot be overstated how important this development was, especially given prevailing ideas at the time. In the end, we will see that Hayek's solution comes in a form that is in many ways remarkably similar (though missing the details of biophysics and anatomy that remained uncovered until the 1990s) to the ideas of AAA, Integrated Information theory, and Predictive Coding and discuss their connections. For Hayek the main questions were:

1. How can a relational network be encoded in the structure and activity of a neural network?¹⁰
2. How are the relations between objects in the outside world learned and encoded (imperfectly) in the neural network of the brain¹¹?

⁹ Psycho-physical parallelism is the idea that there is a one-to-one correspondence between sensory input and the contents of the psyche.

¹⁰ Quote from Hayek: "The question which thus arises for us is how it is possible to construct from the known elements of the neural system a structure which would be capable of performing such discrimination in its response to stimuli as we know our mind in fact to perform." (Hayek 1999).

¹¹ Quote from Hayek: "Our task will be to show how the kind of mechanism which the central nervous system provides may arrange this set of undifferentiated events in an order which possesses the same formal structure as the order of sensory qualities," and "Our task will thus

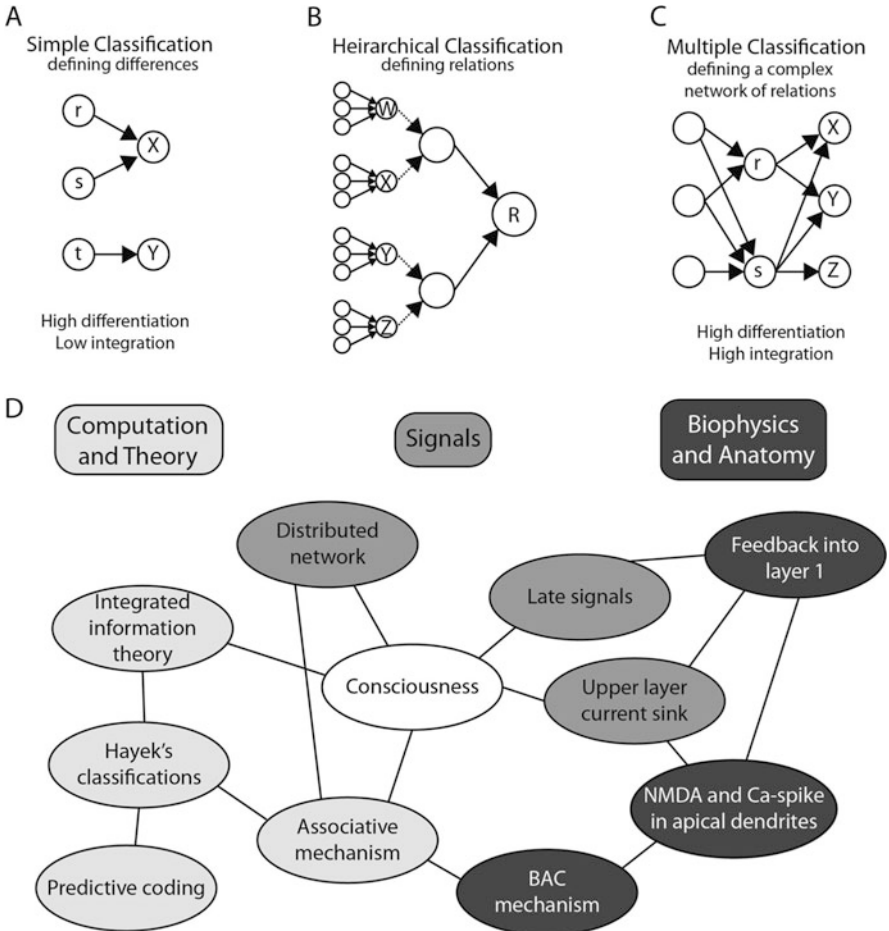


Fig. 8 Hayek’s types of classification and their relationship to Integrated Information Theory. In Hayek’s theory of cortical function, neurons perform a classification function by grouping presynaptic cells that have similar postsynaptic effects together. (a) In simple classification, classes are defined via their different effects on different cells. Here neuron X defines a class {r, s}, because each of that class causes neuron X to fire. Similarly, neuron Y defines a different class {t}. In the conceptual framework of integrated information theory, these “differences that cause a difference” (i.e., the groups {r,s} and {t} each cause different cells to fire) confer the network with high differentiation but not high integration. (b) In hierarchical classification, simple classification occurs in multiple stages. This allows the network to create classes of classes, and, importantly, to classify the *relationships* between different classes. For example, each of neurons W, X, Y, and Z defines a class made up of three cells. The cells postsynaptic to W, X, Y, and Z require two simultaneous inputs to fire, signified by the *dotted lines*. This defines {W&X}, and {Y&Z} as two groups. The neuron R defines a group {W&X,Y&Z}. In this way, the neuron R requires any one of the three cells in groups W and any one of the three cells in group X, or any one of Y and any one of Z, to fire. In this way, the cell R is said to fire to the *relationship* between W and X or to the relationship between Y and Z. Because each of these relationships similarly causes R to fire, these relationships are thus the same. (c) In multiple classifications, neurons can be in multiple classes, and different classes can have overlapping members. In this way, neuron r is in group X and in group Y, and neuron s is in groups X, Y, and Z. In terms of information theory, this type of classification confers the network with integrated information, since neurons r and s have distinct,

The answers to these questions came by positing that a foundational computation the brain performs is classification.¹² Hayek described types of classification of increasing complexity (Fig. 8). *Simple classification* is the sorting of externally different objects into one of a set of different classes by virtue of their differing effects. One example of this is a machine that sorts balls of even diameter into a bin marked A and balls of an odd diameter into a bin marked B. The machine is said to have classified each ball into either group A or B. Simple classification of this sort can describe simple reflexes, which act to group external stimuli by the behaviors that are produced, often by a chain of very few neurons. *Hierarchical classification*¹³ occurs when successive acts of classification occur in successive stages. In this way, the groupings that occur in a previous stage become the objects to be grouped in the next stage. *Multiple classification* allows for stimuli to be in multiple groups at once and also for multiple stimuli to be classified differently than when they occur individually.¹⁴

It is this classification, carried out by the activity of postsynaptic neurons (as a function of presynaptic activity and the structure of anatomical connections), that builds up a system of relations. Here, we already see a conceptual overlap with some modern ideas. For instance, Buzsáki's (2010) *reader* concept is a framework for defining cell assemblies by virtue of postsynaptic effects (e.g., by collective effects on reader neurons). Similarly, an important aspect of *integrated information theory*, which will be discussed more below, is the defining of *causal groups as differences that make a difference*, in other words, defined by their causal postsynaptic effects (Oizumi et al. 2014). There are even mathematical theories of computation in dynamical systems, which have not been created or even used in thinking about neural systems, that use the same conceptual idea, such as epsilon machine reconstruction (Crutchfield 1994), and could potentially be used to analyze network function.

be to show how these undifferentiated individual impulses or groups of impulses may obtain such a position in a system of relations to each other that in their functional significance they will resemble on each other in a manner which corresponds strictly to the relations between the sensory qualities which are evoked by them."

¹² Quote from Hayek: "All the different events which whenever they occur produce the same effect will be said to be events of the same class, and the fact that every one of them produces the same effect will be the sole criterion which makes them members of the same class."

¹³ Hayek does not use the term hierarchical in his description and instead just treats it as a more complicated form of multiple classification.

¹⁴ This classification may thus be 'multiple' in more than one respect. Not only may each individual event belong to more than one class, but it may also contribute to produce different responses of the machine if and only if it occurs in combination with certain other events.



Fig. 8 (continued) but semi-overlapping, causal effects. Thus the network has "differences that cause a difference" but also causal dependencies. **(d)** A conceptual network of the connections between different aspects of biophysics, signals, and theory

In the simplest case of classification, two neurons that individually cause the same postsynaptic effect are seen by the network as being equivalent, that is, as being in one class. Thus, the position of these two neurons in the entire system of relationships is the same. Different neurons will in general have varying degrees of overlap in their postsynaptic effects, making it possible to talk about varying degrees of similarity with respect to their position in the system of relations. In this way, Hayek spoke of the postsynaptic activity *representing* the common attributes of presynaptic impulses that bring about that postsynaptic effect, though he preferred to say that the postsynaptic activity *constitutes* the attribute, rather than *represents* it. This was to make the ontological point that these neural systems *are* what the common attributes actually are and that they do not exist outside of the material actions of the neural network. In other words, the contents of consciousness have a one-to-one correspondence not only with the activity of neurons but also in the structure of the network in which that activity exists. Importantly, this theory differed radically from contemporaneous theories where the qualitative aspects of the mind were somehow attached to the properties of electrical signals themselves. Here, instead, we see the beginnings of an understanding of the psyche that has at its core relations and information: “it is thus the position of the individual impulse or group of impulses in the whole system of connections which gives it its distinctive quality.” (Hayek 1999).

Indeed, it is important to point out that there are two separable aspects of this scheme. The first is the (simple) classification of different signals by their differing effects (“to respond differently to different impulses”). In this way, if each of a group of cells causes the firing of a postsynaptic cell A, and each of a different group of cells causes the firing of a different cell B, then the network has classified these groups of cells into two distinct classes. This alone, however, does not make up a system of relations, because so far we have only described distinct attributes, A and B, with no real relationship between them. The second aspect is then that of putting those attributes in a relationship with one another. This is where multiple classification comes in. By way of example, this process occurs when a postsynaptic cell requires the concurrent input of any of a member of class A alongside any of a member of class B, or the concurrent input of any member of class C and any member of class D. In such a case, we can say that the postsynaptic cell responds to the relationship between A and B, which is the same relationship as between C and D.

These two processes have been put to quantitative work in a modern theory of consciousness, called integrated information theory (IIT), proposed by Giulio Tononi (2008). We will not describe the theory in all of its conceptual and mathematical detail here. For our purposes, it is important to point out the conceptual overlap with Hayek’s ideas of classification, even though the two theories start from a very different set of considerations. The two concepts necessary for Hayek’s scheme to set up a network of relations, that of setting up distinct attributes by virtue of them having distinct postsynaptic effects and that of relating these attributes to each other by virtue of their overlapping (classifying classes) and diverging (being in multiple classes at once) inputs onto postsynaptic cells, can

be conceptually reformulated into the language of information theory. In an information theoretic framework, the setting up of distinct postsynaptic effects (simple classification) confers a high entropy to the network, and thus a high informational content (information here can be estimated as the negative logarithm of the number of different potential states of the system). On the other hand, this information needs to somehow be put in a relational network. This is done by the cooperative effects of different classes both postsynaptically and on each other (multiple classification). This informational dependency is called integration in IIT. Importantly, for a system to have both high information content and high integration, and thus high integrated information, the system must simultaneously have dependencies between different attributes to put them in a relation with one another, but not so much dependency as to erase distinctions between different attributes. Because of a set of principled reasons, IIT posits that systems with high integrated information are conscious.

The exact physiological underpinnings of the thalamocortical system that might give the brain high amounts of integrated information, and thus consciousness, are still illusive. However, by considering the similarities between integrated information and classification, a way forward is seen whereby specific network and physiological structures are found to be plausible candidates. In particular, we will find that, by considering increasingly complex structures of classification (and by extension increasingly complex amounts of integrated information), a hierarchical network of feedforward and feedback interactions can work as the substrate for the representation of conscious percepts.

After the general description of classification and how it can be used to set up a series of relationships, Hayek goes on to find implications for this idea in terms of the structure of the cortex and how it might act to build representations. He begins by considering the simplest of automatic reflexes, which performs a simple classification by grouping sensory inputs by the movements they produce. The evolution of the brain led to these pathways of these reflexes, often carried out by a small number of nerve cells from the periphery to the afferent, branching off and sending axons to higher areas of the brain.¹⁵ This allows the brain to receive information about both the state of the periphery and the actions that the organism is about to take. Unlike pure afferent information, information in the higher centers is available to be used for multiple classification, which can eventually send out motor commands.

Hayek posited a hierarchical scheme whereby the cortex would perform classification in successive layers and could even perform classifications on relationships

¹⁵ It is in this idea, which is the main focus of Chapter 4 in Hayek's book (1999), where Hayek posits a potential use for axons that send the same information to the spinal cord and back to within the cortex. Hayek talks of how there is no evidence for such axons; however, we now know that layer 5b pyramidal neurons have axons that split, sending the same information directly to the spinal cord and to relay cells in the thalamus that feed back into the cortex. The implications of this process has been put into a theory of thalamocortical function, with many parallels to the ideas of Hayek, described by Sherman and Guillery (2002).

themselves, thus providing a highly complex and structured substrate for the psyche. As classifications continue on up the hierarchy, classes become more general and abstracted (classes of classes of classes, and classes of relations between classes, etc.). In the case of the evolution of more complicated control of motor responses, the higher levels can thus act to represent and control more general groups or motor commands. Importantly, sensory input comes into an already active network and thus interacts not only with the anatomical structure of the network but with the activity already present in the network. Hayek describes the type of information processing that feedforward and feedback connections might serve in such a case:

The position of the highest centres [of the brain] in this respect is somewhat like that of the commander of an army (or the head of any other hierarchical organization), who knows that his subordinates will respond to various events in a particular manner, and who will often recognize the character of what has happened as much from the response of his subordinates as from direct observation. It will also be similar in the sense that, so long as the decision taken by his subordinates in the light of their limited but perhaps more detailed observation seems appropriate in view of his more comprehensive knowledge, he will not need to interfere; and that only if something known only to him but not to his subordinates makes those normal responses inappropriate will he have to overrule their decisions by issuing special orders.

In this way, certain cells (or groups of cells) in the brain act by comparing their knowledge with what they receive from sensorium, only interfering in the network when there is a mismatch. A framework for neural computation, called *predictive coding*, is the mathematical description of such a process. The predictive coding framework posits that the brain uses an efficient coding scheme to represent the external world. In particular, this idea posits that natural redundancies in the external environment acting on the sensory apparatus are not explicitly represented in the brain, and instead what is represented is the deviation of the sensory drive from what is predicted. Rao and Ballard (1999) have used this idea to explain the tuning properties of cells in the retina, LGN, and V1. Importantly, this framework puts an emphasis on efficient coding in the brain, something that Hayek did not consider. Despite this, we will see that the biophysical mechanism in which feedforward and feedback signals interact to represent sensory perceptions is conceptually consistent with the predicting coding framework.

In the parlance of predictive coding, feedback signals, from higher to lower levels in the hierarchy, convey predictions of the activity of the lower levels to which they project, that is, predictions of general classes of motor commands given the sensory input. In turn, cells compare predictions with information from lower levels and send error signals forward in the hierarchy. In this way the predictions are continually refined. The diction here becomes conceptually important. A restatement of the processes of refining predictions via error signals representing the comparison of prediction and feedforward sensory driven information puts the ideas regarding network level computation discussed earlier in the chapter squarely in the framework of predictive coding. Indeed, a *comparison* is biophysically nothing more than the local *integration* of feedforward and feedback signals that

occur in a single pyramidal neuron (and influenced by the surrounding local circuit). A *prediction* is the *feedback activity* that predisposes specific neurons in lower cortical areas to varying degrees of activity (or inactivity). An *error* signal is then the result of the integration of feedforward and feedback signals, which are then broadcast to higher areas of the hierarchy. Feedback activity, and its robust control over the output of pyramidal neurons via dendritic nonlinearities (NMDA and Ca-spikes), serves here as a physiological mechanism in which these kinds of computations might be carried out in the brain. Importantly, NMDA and Ca-spikes are physical mechanisms that can be both monitored and exquisitely manipulated in experiment and thus provide a way to test hypotheses about how the cortex implements predicting coding.

In thinking of the further evolution of cortex, Hayek posits that there is fundamentally no difference between the increased control of more complicated motor responses (for instance, being able to account for context) and the representation of complex percepts. Indeed, the addition of more and more layers alongside more complicated forms of classification in the network allows for the network to form a map of the outside world. In this way the role of convergent fibers to higher levels of the hierarchy confers the binding of different attributes into more abstract attributes, whereas divergence confers the distribution of common attributes to different categories (Hayek 1999; Fuster 2003). *Associations* between different attributes are given by connections that predispose but do not on their own elicit activity in a postsynaptic cell or group of cells. Associations of this type can occur in any direction of the hierarchy. An important aspect of such a distributed network is that higher levels of the hierarchy can, through feedback connectivity to lower cortical areas, act to predispose certain cells in the lower areas to fire. Hayek describes the process through which multiple categorization and multiple associations interact to create a dynamic and ongoing *selection* of categories at multiple levels of the hierarchy:

The different associations attaching to individual impulses... will often not only not be convergent but even conflicting; and not all the representations which will form part of the following [ie. postsynaptic effects] of the elements of the complete situation will be capable of simultaneous realization, or would produce a significant new pattern if they did. Since from each element of the structure of connected fibers impulses can pass in a great variety of directions, the initial stream of impulses would merely diffuse and dissipate itself if the overlapping of the following [ie postsynaptic effects] did not determine a selection of some among the many potential paths on which they might travel.

This type of selection, which occurs on account of multiple associations interacting with each other, is consistent with the network level computation that follows from the single-cell biophysics of pyramidal neurons discussed. The main single cell computation in AAA is that of coincident detection or, more generally stated, of integration from multiple axonal pathways. Individually, these pathways bias, but do not cause, the neuron to fire a burst of high-frequency action potentials. The pyramidal neuron thus acts as a highly nonlinear classifier of the thousands of excitatory and inhibitory inputs (and in reality even neuromodulatory inputs) that impinge on the basal and tuft dendrites. These classifying neurons then send their

own long-range axons to many cells in far away areas, establishing hierarchical and multiple classification. In direct analogy to what Hayek discussed, it is the collective action of this process that works to select which pyramidal neurons are active in different areas of the cortex and which act to form the bound representation of percepts in the brain.

The ideas presented in this section are all active areas of research. The connections between these topics (Fig. 8d) range from scientific fact (e.g., NMDA and Ca-spikes) to plausible speculation (the connection between the single cell BAC mechanism and network level binding), or are even philosophical in nature (the relationship between consciousness and binding). In the coming decade, it will be important to establish exactly where, in both mathematical and physiological foundations, these ideas overlap and differ. At the very least, Hayek's stream of thought suggests that there are connections waiting to be uncovered. Ultimately, understanding the cortical network implications of single cell and local network computation would be made easier if a more direct connection between ideas like AAA, which explicitly take into account physiological and anatomical details of the type that are experimentally measurable and readily manipulated, and the more theoretical ideas of network computation like predictive coding and IIT was better understood.

Acknowledgments We would like to thank Nathan Faivre for invaluable comments and discussions on the manuscript. We would also like to thank Christof Koch and Gyorgy Buzsaki for providing us a venue to report these thoughts and considerations. Finally, we both are thankful to the foundations that support our work: the G. Harold and Leila Y. Mathers foundation, the National Institutes of Health, the National Science Foundation, the Swiss National Science Foundation, the Human Frontier Sciences Programme, the Whitaker International Program and the Paul and Jodie Allen foundation.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Akemann W, Mutoh H, Perron A, Rossier J, Knöpfel T (2010) Imaging brain electric signals with genetically targeted voltage-sensitive fluorescent proteins. *Nat Methods* 7:643–649
- Anastassiou CA, Perin R, Buzsaki G, Markram H, Koch C (2015) Cell-type- and activity-dependent extracellular correlates of intracellular spiking. *J Neurophysiol* 114(1):608–623. doi:10.1152/jn.00628.2014

- Aru J, Axmacher N, Do Lam AT, Fell J, Elger CE, Singer W, Melloni L (2012) Local category-specific gamma band responses in the visual cortex do not reflect conscious perception. *J Neurosci* 32:14909–14914
- Baars BJ (2005) Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog Brain Res* 150:45–53
- Barthó P, Hirase H, Monconduit L, Zugaro M, Harris KD, Buzsáki G (2004) Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *J Neurophysiol* 92:600–608
- Bazelot M, Dinocourt C, Cohen I, Miles R (2010) Unitary inhibitory field potentials in the CA3 region of rat hippocampus. *J Physiol* 588:2077–2090
- Bédard C, Destexhe A (2009) Macroscopic models of local field potentials and the apparent 1/f noise in brain activity. *Biophys J* 96:2589–2603
- Bédard C, Kröger H, Destexhe A (2004) Modeling extracellular field potentials and the frequency-filtering properties of extracellular space. *Biophys J* 86:1829–1842
- Belluscio MA, Mizuseki K, Schmidt R, Kempter R, Buzsáki G (2012) Cross-frequency phase-phase coupling between θ and γ oscillations in the hippocampus. *J Neurosci* 32:423–435
- Blake R, Fox R (1974) Binocular rivalry suppression: insensitive to spatial frequency and orientation change. *Vision Res* 14:687–692
- Breitmeyer BG, Ogmen H (2000) Recent models and findings in visual backward masking: a comparison, review, and update. *Percept Psychophys* 62:1572–1595
- Brombas A, Fletcher LN, Williams SR (2014) Activity-dependent modulation of layer 1 inhibitory neocortical circuits by acetylcholine. *J Neurosci* 34:1932–1941
- Butos WN, Koppl RG (2007) Does the sensory order have a useful economic future? *Cogn Econ Adv Austrian Econ* 9:19–50
- Buzsáki G (2004) Large-scale recording of neuronal ensembles. *Nat Neurosci* 7:446–451
- Buzsáki G (2010) Neural syntax: cell assemblies, synapse ensembles, and readers. *Neuron* 68:362–385
- Buzsáki G, Mizuseki K (2014) The log-dynamic brain: how skewed distributions affect network operations. *Nat Rev Neurosci* 15:264–278
- Buzsáki G, Penttonen M, Nádasdy A, Bragin A (1996) Pattern and inhibition-dependent invasion of pyramidal cell dendrites by fast spikes in the hippocampus in vivo. *Proc Natl Acad Sci USA* 93:9921–9925
- Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat Rev Neurosci* 13:407–420
- Caldwell B (2004) Some reflections on FA Hayek's the sensory order. *J Bioecon* 6:239–254
- Carandini M, Ringach DL (1997) Predictions of a recurrent model of orientation selectivity. *Vision Res* 37:3061–3071
- Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, Casali KR, Casarotto S, Bruno MA, Laureys S, Tononi G, Massimini M (2013) A theoretically based index of consciousness independent of sensory processing and behavior. *Sci Transl Med* 5(198):198ra105. doi:10.1126/scitranslmed.3006294
- Cauler LJ, Kulics AT (1988) A comparison of awake and sleeping cortical states by analysis of the somatosensory-evoked response of postcentral area 1 in Rhesus monkey. *Exp Brain Res* 72:584–592
- Colgin LL, Denninger T, Fyhn M, Hafting T, Bonnevie T, Jensen O, Moser MB, Moser EI (2009) Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature* 462:353–357
- Connors BW, Benardo LS, Prince DA (1983) Coupling between neurons of the developing rat neocortex. *J Neurosci* 3:773–782
- Crick F, Koch C (1990) Towards a neurobiological theory of consciousness. In: *Seminars in the neurosciences*. Saunders, pp 263–275. <http://authors.library.caltech.edu/40352/>. Accessed 12 Nov 2015
- Crick F, Koch C (1995) Are we aware of neural activity in primary visual cortex? *Nature* 375:121–123

- Crick F, Koch C (2003) A framework for consciousness. *Nat Neurosci* 6:119–126
- Crutchfield JP (1994) The calculi of emergence: computation, dynamics and induction. *Phys D Nonlinear Phenom* 75:11–54
- Cumming BG, Parker AJ (1997) Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature* 389:280–283
- D’Ambrosio R, Wenzel J, Schwartzkroin PA, McKhann GM, Janigro D (1998) Functional specialization and topographic segregation of hippocampal astrocytes. *J Neurosci* 18:4425–4438
- De Graaf TA, Hsieh P-J, Sack AT (2012) The “correlates” in neural correlates of consciousness. *Neurosci Biobehav Rev* 36:191–197
- De Kock CPJ, Sakmann B (2008) High frequency action potential bursts (≥ 100 Hz) in L2/3 and L5B thick tufted neurons in anaesthetized and awake rat primary somatosensory cortex. *J Physiol* 586(Pt 14):3353–3364
- Dehaene S, Sergent C, Changeux J-P (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc Natl Acad Sci USA* 100:8520–8525
- Dehghani N, Bédard C, Cash SS, Halgren E, Destexhe A (2010) Comparative power spectral analysis of simultaneous electroencephalographic and magnetoencephalographic recordings in humans suggests non-resistive extracellular media. *J Comput Neurosci* 29:405–421
- Denk W, Delaney KR, Gelperin A, Kleinfeld D, Strowbridge BW, Tank DW, Yuste R (1994) Anatomical and functional imaging of neurons using 2-photon laser scanning microscopy. *J Neurosci Methods* 54:151–162
- Dennett DC, Kinsbourne M (1992) Time and the observer: the where and when of consciousness in the brain. *Behav Brain Sci* 15:183–201
- Druckmann S, Banitt Y, Gidon A, Schürmann F, Markram H, Segev I (2007) A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. *Front Neurosci* 1:7–18
- Ebersole JS, Ebersole SM (2010) Combining MEG and EEG source modeling in epilepsy evaluations. *J Clin Neurophysiol* 27:360–371
- Edelman GM (1993) Neural Darwinism: selection and reentrant signaling in higher brain function. *Neuron* 10:115–125
- Einevoll GT, Kayser C, Logothetis NK, Panzeri S (2013) Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nat Rev Neurosci* 14:770–785
- Eliasmith C, Stewart TC, Choo X, Bekolay T, DeWolf T, Tang Y, Rasmussen D (2012) A large-scale model of the functioning brain. *Science* 338:1202–1205
- Elul R (1971) The genesis of the EEG. *Int Rev Neurobiol* 15:227–272
- Fisch L, Privman E, Ramot M, Harel M, Nir Y, Kipervasser S, Andelman F, Neufeld MY, Kramer U, Fried I, Malach R (2009) Neural “ignition”: enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* 64:562–574
- Flourens P (1842) *Recherches expérimentales sur les propriétés et les fonctions du système nerveux dans les animaux vertébrés*. Ballière. https://books.google.com/books?hl=en&lr=&id=_WRZW_d4R0IC&oi=fnd&pg=PA1&dq=florens&ots=VVspXxOcmI&sig=UEZm6vPd4Sy0rjO35u0YRTnuquU. Accessed 12 Nov 2015
- Friston K (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11:127–138
- Fuster JM (2003) *Cortex and mind: unifying cognition*. Oxford University Press, New York, <http://psycnet.apa.org/psycinfo/2002-18891-000>. Accessed 12 Nov 2015
- Fuster JM (2006) The cognit: a network model of cortical representation. *Int J Psychophysiol* 60:125–132
- Gabriel S, Lau RW, Gabriel C (1996) The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz. *Phys Med Biol* 41:2251
- Gaillard R, Dehaene S, Adam C, Clémenceau S, Hasboun D, Baulac M, Cohen L, Naccache L (2009) Converging intracranial markers of conscious access. *PLoS Biol* 7(3):e1000061

- Gawne TJ, Martin JM (2000) Activity of primate V1 cortical neurons during blinks. *J Neurophysiol* 84:2691–2694
- Glickfeld L, Roberts JD, Somogyi P, Scanziani M (2009) Interneurons hyperpolarize pyramidal cells along their entire somatodendritic axis. *Nat Neurosci* 12:21–23
- Gold C, Henze DA, Koch C, Buzsáki G (2006) On the origin of the extracellular action potential waveform: a modeling study. *J Neurophysiol* 95:3113–3128
- Gold C, Girardin CC, Martin KAC, Koch C (2009) High-amplitude positive spikes recorded extracellularly in cat visual cortex. *J Neurophysiol* 102:3340–3351
- Goldstein K (1942) Aftereffects of brain injuries in war: their evaluation and treatment. The application of psychologic methods in the clinic. Grune & Stratton, Oxford, UK, <http://psycnet.apa.org/psycinfo/1943-00160-000>. Accessed 12 Nov 2015
- Goto T, Hatanaka R, Ogawa T et al (2010) An evaluation of the conductivity profile in the somatosensory barrel cortex of Wistar rats. *J Neurophysiol* 104:3388–3412
- Grinvald A, Hildesheim R (2004) VSDI: a new era in functional imaging of cortical dynamics. *Nat Rev Neurosci* 5:874–885
- Haider M, Spong P, Lindsley DB (1964) Attention, vigilance, and cortical evoked-potentials in humans. *Science* 145:180–182
- Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV (1993) Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev Mod Phys* 65:413–497
- Hameroff SR (1994) Quantum coherence in microtubules: a neural basis for emergent consciousness? *J Conscious Stud* 1:91–118
- Harlow JM (1999) Passage of an iron rod through the head. 1848. *J Neuropsychiatr Clin Neurosci* 11:281–283
- Harris KD, Hirase H, Leinekugel X, Henze DA, Buzsáki G (2001) Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. *Neuron* 32:141–149
- Hay E, Hill S, Schürmann F, Markram H, Segev I (2011) Models of neocortical layer 5b pyramidal cells capturing a wide range of dendritic and perisomatic active properties. *PLoS Comput Biol* 7(7):e1002107
- Hayek FA (1991) Contributions to a theory of how consciousness develops. Translated by Grete Heinz. Hoover Institution, Hayek Archives, Box 92
- Hayek FA (1999) The sensory order: an inquiry into the foundations of theoretical psychology. University of Chicago Press, Chicago, IL, https://books.google.com/books?hl=en&lr=&id=UFazm1Xy_j4C&oi=fnd&pg=PR6&dq=The+Sensory+Order:+An+Inquiry+into+the+Foundations+of+Theoretical+Psychology+&ots=8M8XQppbR1&sig=X8dwgbN0lvfxmJklkhsSbTNS9il. Accessed 12 Nov 2015
- Heinen K, Jolij J, Lamme VAF (2005) Figure-ground segregation requires two distinct periods of activity in V1: a transcranial magnetic stimulation study. *Neuroreport* 16:1483–1487
- Henze DA, Borhegyi Z, Csicsvari J, Mamiya A, Harris KD, Buzsáki G (2000) Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *J Neurophysiol* 84:390–400
- Hill S, Tononi G (2005) Modeling sleep and wakefulness in the thalamocortical system. *J Neurophysiol* 93:1671–1698
- Hille B (1992) Ion channels of excitable membranes, 3rd edn. Sinauer Associates, Sunderland, MA
- Hoeltzell PB, Dykes RW (1979) Conductivity in the somatosensory cortex of the cat—evidence for cortical anisotropy. *Brain Res* 177:61–82
- Holt GR (1998) A critical reexamination of some assumptions and implications of cable theory in neurobiology. PhD, California Institute of Technology. <http://resolver.caltech.edu/CaltechETD:etd-09122006-135415>. Accessed 8 Apr 2015
- Holt GR, Koch C (1999) Electrical interactions via the extracellular potential near cell bodies. *J Comput Neurosci* 6:169–184

- Horwitz S (2000) From the sensory order to the liberal order: Hayek's non-rationalist liberalism. *Rev Austrian Econ* 13:23–40
- Hu H, Gan J, Jonas P (2014) Interneurons. Fast-spiking, parvalbumin⁺ GABAergic interneurons: from cellular design to microcircuit function. *Science* 345:1255–1263
- Hubel DH (1982) Cortical neurobiology: a slanted historical perspective. *Annu Rev Neurosci* 5:363–370
- Imas OA, Ropella KM, Ward BD, Wood JD, Hudetz AG (2005) Volatile anesthetics disrupt frontal-posterior recurrent information transfer at gamma frequencies in rat. *Neurosci Lett* 387:145–150
- Jarsky T, Roxin A, Kath WL, Spruston N (2005) Conditional dendritic spike propagation following distal synaptic activation of hippocampal CA1 pyramidal neurons. *Nat Neurosci* 8:1667–1676
- Jiang X, Wang G, Lee AJ, Stornetta RL, Zhu JJ (2013) The organization of two new cortical interneuronal circuits. *Nat Neurosci* 16:210–218
- Juan CH, Walsh V (2003) Feedback to V1: a reverse hierarchy in vision. *Exp Brain Res* 150:259–263
- Katzner S, Nauhaus I, Benucci A, Bonin V, Ringach DL, Carandini M (2009) Local origin of field potentials in visual cortex. *Neuron* 61:35–41
- Khodagholy D, Gelineau JN, Thesen T, Doyle W, Devinsky O, Malliaras GG, Buzsáki G (2015) NeuroGrid: recording action potentials from the surface of the brain. *Nat Neurosci* 18:310–315
- King J-R, Sitt JD, Faugeras F, Rohaut B, Karoui IEL, Cohen L, Naccache L, Dehaene S (2013) Information sharing in the brain indexes consciousness in noncommunicative patients. *Curr Biol* 23:1914–1919
- Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 27:712–719
- Koch C (2004) *Biophysics of computation: information processing in single neurons*. Oxford University Press, Oxford
- Koivisto M, Revonsuo A (2003) An ERP study of change detection, change blindness, and visual awareness. *Psychophysiology* 40(3):423–429
- Koivisto M, Revonsuo A (2007) Electrophysiological correlates of visual consciousness and selective attention. *Neuroreport* 18(8):753–756
- Koivisto M, Revonsuo A (2010) Event-related brain potential correlates of visual awareness. Special section: developmental determinants of sensitivity and resistance to stress: a tribute to Seymour “Gig” Levine. *Neurosci Biobehav Rev* 34:922–934
- Koivisto M, Lähteenmäki M, Sørensen TA, Vangkilde S, Overgaard M, Revonsuo A (2008) The earliest electrophysiological correlate of visual awareness? *Brain Cogn* 66:91–103
- Kreiman G, Hung CP, Kraskov A, Quiroga RQ, Poggio T, DiCarlo JJ (2006) Object selectivity of local field potentials and spikes in the Macaque inferior temporal cortex. *Neuron* 49:433–445
- Kulics AT, Cauller LJ (1986) Cerebral cortical somatosensory evoked responses, multiple unit activity and current source-densities: their interrelationships and significance to somatic sensation as revealed by stimulation of the awake monkey's hand. *Exp Brain Res* 62:46–60
- Kulics AT, Cauller LJ (1989) Multielectrode exploration of somatosensory cortex function in the awake monkey. Sensory processing in the mammalian brain: neural substrates and experimental strategies. *CNUP Neurosci Rev* 85–115
- Lamme VA (2001) Blindsight: the role of feedforward and feedback corticocortical connections. *Acta Psychol* 107:209–228
- Lamme VAF (2006) Towards a true neural stance on consciousness. *Trends Cogn Sci* 10:494–501
- Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci* 23:571–579
- Lamme VAF, Zipser K, Spekreijse H (1998) Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proc Natl Acad Sci USA* 95:3263–3268
- Larkum M (2013) A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci* 36:141–151

- Larkum ME, Zhu JJ, Sakmann B (1999) A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature* 398:338–341
- Larkum ME, Nevian T, Sandler M, Polsky A, Schiller J (2009) Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. *Science* 325:756–760
- Lashley KS (1929) Brain mechanisms and intelligence: a quantitative study of injuries to the brain. <http://psycnet.apa.org/psycinfo/2004-16230-000/>. Accessed 12 Nov 2015
- Lashley KS (1950) In search of the engram. <http://gureckislab.org/courses/fall13/leammem/papers/Lashley1950.pdf>. Accessed 12 Nov 2015
- Lee JH, Durand R, Gradinaru V, Zhang F, Goshen I, Kim DS, Fenno LE, Ramakrishnan C, Deisseroth K (2010) Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. *Nature* 465:788–792
- Lindén H, Pettersen KH, Einevoll GT (2010) Intrinsic dendritic filtering gives low-pass power spectra of local field potentials. *J Comput Neurosci* 29:423–444
- Lindén H, Tetzlaff T, Potjans TC, Pettersen KH, Grün S, Diesmann M, Einevoll GT (2011) Modeling the spatial reach of the LFP. *Neuron* 72:859–872
- Lisman JE (1997) Bursts as a unit of neural information: making unreliable synapses reliable. *Trends Neurosci* 20:38–43
- Liu J, Newsome WT (2006) Local field potential in cortical area MT: stimulus tuning and behavioral correlations. *J Neurosci* 26:7779–7790
- Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. *Annu Rev Physiol* 66:735–769
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157
- Logothetis NK, Kayser C, Oeltermann A (2007) In vivo measurement of cortical impedance spectrum in monkeys: implications for signal propagation. *Neuron* 55(5):809–823
- Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9:1432–1438
- Markram H, Anirudh D, Gupta A, Uziel A, Wang Y, Tsodyks M (1998) Information processing with frequency-dependent synaptic connections. *Neurobiol Learn Mem* 70:101–112
- Markram H, Muller E, Ramaswamy S, Reimann M, King JG (2015) Reconstruction and simulation of neocortical microcircuitry. *Cell* 163:456–492
- Massimini M, Ferrarelli F, Huber R, Esser SK, Singh H, Tononi G (2005) Breakdown of cortical effective connectivity during sleep. *Science* 309:2228–2232
- McFadden J (2002) The conscious electromagnetic information (cemi) field theory: the hard problem made easy? *J Conscious Stud* 9:45–60
- Melloni L, Schwiedrzik CM, Müller N, Rodriguez E, Singer W (2011) Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. *J Neurosci* 31:1386–1396
- Mitzdorf U (1985) Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Am Physiol Soc*. <http://physrev.physiology.org/content/physrev/65/1/37.full.pdf>. Accessed 14 Nov 2015
- Nicholson C, Freeman JA (1975) Theory of current source-density analysis and determination of conductivity tensor for anuran cerebellum. *J Neurophysiol* 38:356–368
- Niedermeyer E, Lopes da Silva FH (2005) Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins, Baltimore, MD
- Nir Y, Fisch L, Mukamel R, Gelbard-Sagiv H, Arieli A, Fried I, Malach R (2007) Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. *Curr Biol* 17:1275–1285
- Nowak LG, Bullier J (1997) The timing of information transfer in the visual system in extrastriate cortex in primates. *Springer*, pp 205–241. http://link.springer.com/chapter/10.1007/978-1-4757-9625-4_5. Accessed 12 Nov 2015
- Nunez PL, Srinivasan R (2006) Electric fields of the brain: the neurophysics of EEG. Oxford University Press, Oxford

- Oizumi M, Albantakis L, Tononi G (2014) From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. <http://dx.plos.org/10.1371/journal.pcbi.1003588>. Accessed 12 Nov 2015
- Palmer LM, Schulz JM, Murphy SC, Ledergerber D, Murayama M, Larkum ME (2012) The cellular basis of GABAB-mediated interhemispheric inhibition. *Science* 335:989–993
- Palmer LM, Shai AS, Reeve JE, Anderson HL, Paulsen O, Larkum ME (2014) NMDA spikes enhance action potential generation during sensory input. *Nat Neurosci* 17:383–390
- Pascual-Leone A, Walsh V (2001) Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292:510–512
- Pearce JMS (2009) Marie-Jean-Pierre Flourens (1794–1867) and cortical localization. *Eur Neurol* 61:311–314
- Perea G, Araque A (2007) Astrocytes potentiate transmitter release at single hippocampal synapses. *Science* 317:1083–1086
- Pereda AE (2014) Electrical synapses and their functional interactions with chemical synapses. *Nat Rev Neurosci* 15:250–263
- Pérez-Garci E, Gassmann M, Bettler B, Larkum ME (2006) The GABA B1b isoform mediates long-lasting inhibition of dendritic Ca²⁺ spikes in layer 5 somatosensory pyramidal neurons. *Neuron* 50:603–616
- Pérez-Garci E, Larkum ME, Nevian T (2013) Inhibition of dendritic Ca²⁺ spikes by GABAB receptors in cortical pyramidal neurons is mediated by a direct Gi/o- $\beta\gamma$ -subunit interaction with Cav1 channels. *J Physiol* 591:1599–1612
- Pettersen KH, Einevoll GT (2008) Amplitude variability and extracellular low-pass filtering of neuronal spikes. *Biophys J* 94:784–802
- Pitts W (1952) Investigations on synaptic transmission. In: *Cybernetics - Transactions of the ninth conference of the Josiah Macy Foundation, New York*, pp 159–162
- Pockett S (2012) The electromagnetic field theory of consciousness a testable hypothesis about the characteristics of conscious as opposed to non-conscious fields. *J Conscious Stud* 19:191–223
- Railo H, Koivisto M, Revonsuo A (2011) Tracking the processes behind conscious perception: a review of event-related potential correlates of visual consciousness. *Conscious Cogn* 20:972–983
- Rall W, Shepherd GM (1968) Theoretical reconstruction of field potentials and dendrodendritic synaptic interactions in olfactory bulb. *J Neurophysiol* 31:884–915
- Ranck JB (1973) Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats. I. Behavioral correlates and firing repertoires. *Exp Neurol* 41:461–531
- Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87
- Ray S, Maunsell JHR (2011) Different origins of gamma rhythm and high-gamma activity in Macaque visual cortex. *PLoS Biol* 9:e1000610
- Rees G, Kreiman G, Koch C (2002) Neural correlates of consciousness in humans. *Nat Rev Neurosci* 3:261–270
- Reimann MW, Anastassiou CA, Perin R, Hill SL, Makram H, Koch C (2013) A biophysically detailed model of neocortical local field potentials predicts the critical role of active membrane currents. *Neuron* 79:375–390
- Revonsuo A (1999) Binding and the phenomenal unity of consciousness. *Conscious Cogn* 8:173–185
- Ringach DL, Hawken MJ, Shapley R (1997) Dynamics of orientation tuning in Macaque primary visual cortex. *Nature* 387:281–284
- Rosenfalck P (1969) Intra- and extracellular potential fields of active nerve and muscle fibres. A physico-mathematical analysis of different models. *Acta Physiol Scand Suppl* 321:1–168
- Royer S, Zemelman BV, Losonczy A, Kim J, Chance F, Magee JC, Buzsáki G (2012) Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. *Nat Neurosci* 15:769–775

- Sachidhanandam S, Sreenivasan V, Kyriakatos A, Kremer Y, Petersen CCH (2013) Membrane potential correlates of sensory perception in mouse barrel cortex. *Nat Neurosci* 16:1671–1677
- Salti M, Monto S, Charles L, King JR, Parkkonen L, Dehaene S (2015) Distinct cortical codes and temporal dynamics for conscious and unconscious percepts. *eLife* e05652
- Sarasso S, Boly M, Napolitani M, Gosseries O, Charland-Verville V, Casarotto S, Rosanova M, Girardi Casali A, Brichant JF, Boveroux P, Rex S, Tononi G, Laureys S, Massimini M. Consciousness and complexity during unresponsiveness induced by propofol, xenon, and ketamine. *Curr Biol* 0(0). <http://www.cell.com/article/S0960982215012427/abstract>. Accessed 20 Nov 2015
- Schölvinck ML, Maier A, Ye FQ, Duyn JH, Leopold DA (2010) Neural basis of global resting-state fMRI activity. *Proc Natl Acad Sci USA* 107:10238–10243
- Schomburg EW, Anastassiou CA, Buzsáki G, Koch C (2012) The spiking component of oscillatory extracellular potentials in the rat hippocampus. *J Neurosci* 32:11798–11811
- Schomburg EW, Fernández-Ruiz A, Mizuseki K, Berényi A, Anastassiou CA, Koch C, Buzsáki G (2014) Theta phase segregation of input-specific gamma patterns in entorhinal-hippocampal networks. *Neuron* 84:470–485
- Shai AS, Koch C, Anastassiou CA (2014) Spike-timing control by dendritic plateau potentials in the presence of synaptic barrages. *Front Comput Neurosci* 8:89, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4132263/>. Accessed 12 Nov 2015
- Shai AS, Anastassiou CA, Larkum ME, Koch C (2015) Physiology of layer 5 pyramidal neurons in mouse primary visual cortex: coincidence detection through bursting. *PLoS Comput Biol* 11: e1004090
- Sherman SM, Guillery RW (2002) The role of the thalamus in the flow of information to the cortex. *Philos Trans R Soc Lond B Biol Sci* 357:1695–1708
- Sherman SM, Guillery RW (2011) Distinct functions for direct and transthalamic corticocortical connections. *J Neurophysiol* 106:1068–1077
- Siegel MS, Isacoff EY (1997) A genetically encoded optical probe of membrane voltage. *Neuron* 19:735–741
- Silverstein BH, Snodgrass M, Shevrin H, Kushwaha R (2015) P3b, consciousness, and complex unconscious processing. *Cortex* 73:216–227
- Spong P, Haider M, Lindsley DB (1965) Selective attentiveness and cortical evoked responses to visual and auditory stimuli. *Science* 148:395–397
- Sugase Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400:869–873
- Szabadi J, Lorincz A, Tamás G (2001) β and γ frequency synchronization by dendritic GABAergic synapses and gap junctions in a network of cortical interneurons. *J Neurosci* 21:5824–5831
- Taxidis J, Anastassiou CA, Diba K, Koch C (2015) Local field potentials encode place cell ensemble activation during hippocampal sharp wave ripples. *Neuron* 87:590–604
- Tononi G (2008) Consciousness as integrated information: a provisional manifesto. *Biol Bull* 215:2016–2242
- Traub RD, Kopell N, Bibbig A, Traub RD, Kopell N, Bibbig A, Buhl EH, LeBeau FEN, Whittington MA (2001) Gap junctions between interneuron dendrites can enhance synchrony of gamma oscillations in distributed networks. *J Neurosci* 21:9478–9486
- Trayanova N, Henriquez CS (1991) Modification of a cylindrical bidomain model for cardiac tissue. *Math Biosci* 104:59–72
- Tsodyks MV, Markram H (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc Natl Acad Sci USA* 94:719–723
- Tsodyks M, Pawelzik K, Markram H (1998) Neural networks with dynamic synapses. *Neural Comput* 10:821–835
- Tsuchiya N, Wilke M, Frässle S, Lamme VAF (2015) No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn Sci*. doi:10.1016/j.tics.2015.10.002

- Volterra A, Meldolesi J (2005) Astrocytes, from brain glue to communication elements: the revolution continues. *Nat Rev Neurosci* 6:626–640
- Wang X, Lou N, Xu Q, Tian GF, Peng WG, Han X, Kang J, Takano T, Nedergaard M (2006) Astrocytic Ca²⁺ signaling evoked by sensory stimulation in vivo. *Nat Neurosci* 9:816–823
- Weiskrantz L (1986) *Blindsight: a case study and implications*. Oxford University Press, Oxford
- Whittingstall K, Logothetis NK (2009) Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. *Neuron* 64:281–289
- Xing D, Yeh C-I, Shapley RM (2009) Spatial spread of the local field potential and its laminar variation in visual cortex. *J Neurosci* 29:11540–11549
- Yang W, Carrasquillo Y, Hooks BM, Nerbonne JM, Burkhalter A (2013) Distinct balance of excitation and inhibition in an interareal feedforward and feedback circuit of mouse visual cortex. *J Neurosci* 33:17373–17384
- Yuille A, Kersten D (2006) Vision as Bayesian inference: analysis by synthesis? *Trends Cogn Sci* 10:301–308
- Zanos TP, Mineault PJ, Pack CC (2011) Removal of spurious correlations between spikes and local field potentials. *J Neurophysiol* 105:474–486
- Zipser K, Lamme VAF, Schiller PH (1996) Contextual modulation in primary visual cortex. *J Neurosci* 16:7376–7389

Federating and Integrating What We Know About the Brain at All Scales: Computer Science Meets the Clinical Neurosciences

Richard Frackowiak, Anastasia Ailamaki, and Ferath Kherif

Abstract Our everyday professional and personal lives are irrevocably affected by technologies that search and understand the meaning of data, that store and preserve important information, and that automate complex computations through algorithmic abstraction. People increasingly rely on products from computer companies such as Google, Apple, Microsoft and IBM, not to mention their spinoffs, apps, WiFi, iCloud, HTML, smartphones and the like. Countless daily tasks and habits, from shopping to reading, entertainment, learning and the visual arts, have been profoundly altered by this technological revolution. Science has also benefited from this rapid progress in the field of information and computer science and associated technologies (ICT). For example, the tentative confirmation of the existence of the Higgs boson (CMS Collaboration et al. *Phys Lett B* 716:30–61, 2012), made through a combination of heavy industrial development, internet-based scientific communication and collaboration, with data federation, integration, mining and analysis (Rajasekar et al. *iRODS primer: integrated rule-oriented data system. Synthesis lectures on information concepts, retrieval, and services*. Morgan & Claypool, San Rafael, 2010; Chiang et al. *BMC Bioinformatics* 12:361, 2011; Marks. *New Sci* 196:28–29, 2007), has taken our understanding of the structure of inorganic matter to a new level (Hay et al. *The fourth paradigm: data-intensive scientific discovery*. Microsoft, Redmond, WA, 2009). But within this vision of universal progress, there is one anomaly: the relatively poor exploitation and application of new ICT techniques in the context of the clinical neurosciences. A pertinent example is the genetic study of brain diseases and associated bioinformatics methods. Despite a decade of work on clinically well-defined cohorts, disappointment remains among some that genome-wide association studies (GWAS) have not solved many questions of disease causation, especially in psychiatry (Goldstein. *N Engl J Med* 360:1696–1698, 2009). One question is

R. Frackowiak (✉) • F. Kherif

Department of Clinical Neurosciences, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland

e-mail: richard.frackowiak@gmail.com

A. Ailamaki

Department of Computer Science, Ecole Polytechnique Fédérale de Lausanne, 1021 Lausanne, Switzerland

whether we have the appropriate disease categories. Another factor is that gene expression is affected by environmental and endogenous factors, as is protein function in different circumstances (think of the effects of age, developmental stage and nutrition). It is clear that any genetic associations with disease expression are likely to be highly complex. Why then are the world's most powerful super-computers not being deployed with novel algorithms grounded in complexity mathematics to identify biologically homogeneous disease types, or to understand the many interactions that lead to the integrated functions that arise from DNA metabolism, such as cognition? Is it from a lack of appropriate data and methods or are the reasons related to our current clinical scientific culture?

Introduction

Reductionist methods of hypothesis falsification have dominated science in the last two centuries, and rightly so, given the major advances in knowledge about the living and non-living worlds they have afforded. However, there is much evidence to suggest that a uniquely reductionist approach may be blinkered. Indeed it may always have been so—think of Linnaean categorization of the plant kingdom as a scoping exercise prior to a more modern hypothesis-led, genetically based description of plant biology. Darwin had no idea about the physical nature of the hereditary process he so cleverly deduced (he received one of the few original copies of Gregor Mendel's manuscript, but did not read it, judging by the fact that it was found uncut at his death). Yet he catalogued the animal kingdom, discovering hidden patterns that gave rise to his theory about adaptive mechanisms and successful procreation underlying the evolution of species. But do biomedical granting agencies fund work that does not express a firm and clear hypothesis? What modern biomedical grants agency will fund scoping studies involving observation and classification (though again, perhaps that's what GWAS studies are)? Outside epidemiology, such a scenario is difficult to entertain. And in epidemiology, how many studies emerge from the correlative world of univariate statistics, and how many founder on inadequate power?

The most powerful means of examining the spread of influenza epidemics is now achieved by analyzing the geographical spread of incidence posted on Google (Brownstein et al. 2008; <http://www.google.org/flutrends/>). This is a real-time example of the interconnected power of global computing; crowd sourcing is another (Brabham 2008). How organic matter self-organizes across spatial and temporal scales to produce the diversity of living, reproducing, adaptive creatures and their nervous systems is a question that is slowly being addressed with a new methodological agenda. The complexity of the human brain demands modern methods that address, describe and quantify interactions in large integrated systems.

Clinicians need to take note of this trend, both in terms of the science and art of medicine and also in any effort to rapidly identify and develop effective treatments.

Syndromic Diagnosis

What is the challenge? Firstly, the clinical-pathological paradigm of the last century and a half, attributed to Broca in the clinical neurosciences, has reached the limits of its usefulness. Syndromes, composed of groups of symptoms narrated by patients with varying degrees of cognitive impairment, or by their relatives, to individual practitioners, overlap too much to remain useful as a basis for the precise diagnosis of brain diseases. This is not a new insight, as demonstrated by the variability in presentation of diseases such as syphilis and diabetes mellitus, but it is an increasingly pertinent one. Recently it has been reported that the five major classes of psychiatric illness share a similar set of associated genes that predispose not to one or other class but to mental illness in general (Cross-Disorder Group of the Psychiatric Genomics Consortium 2004). The spinocerebellar ataxias are associated with well over 20 dominant, often partially penetrant, mutations, each of which generates a similar pattern of clinical features, at times causing diagnostic confusion (Schöls et al. 2012). The dementia syndrome is caused by a range of pathological mechanisms, a few of which are genetically determined, the vast majority of which are of unknown aetiology, to the extent that the diagnosis of Alzheimer's disease (AD) is wrong in the best centers about 30 % of the time, if post mortem features are used to define disease (Beach et al. 2012). Longitudinal syndromic studies demonstrate that even diagnoses of "pure" syndromes fail to remain applicable through life, and correlation with post mortem features is poor if not random (Kertesz et al. 2005). Finally, the same single genetic mutation can present with a variety of syndromes. A simple example is that of Huntington's disease, where a behavioral or psychiatric presentation is recognized, as are presentations with movement disorders or gait abnormalities. Though the phenomenon of generational anticipation in male presentation of the disease is associated with the length of CAG repeats in the huntingtin gene, it is not understood how this happens. In short, there is a pressing need to move from an observational and simple correlational approach to clinical neuroscience to one that is mechanistic and multifactorial.

A Theory of the Brain

That is easier said than done, for a simple reason. Unlike the materials sciences, where there is a clear if still often approximate (except at the quantum level) understanding of the organization of inorganic matter across spatial and temporal scales, no such theory of living matter is available. However, this is not an intractable problem with infinite degrees of freedom, as some have suggested.

The building block of organic matter, DNA, is composed of a limited set of highly specific base pairs. We have a good understanding of how transcription to RNA and translation to proteins occur, and what mechanisms control these processes. The human genome is known and much if not all of the variation in it has been catalogued. Much of it consists of (mysterious) non-coding sequences. That takes care of a lot of degrees of freedom and sets parameters on how life itself emerges, as well as cognition, emotion, perception and action. The rules that determine mechanistic interactions at these basic levels are constantly being discovered but remain unconnected without a global theory of brain organization from the lowest to the highest levels: from base pairs, to genes, to functional and structural proteins, to neurons and glia, to cortical columns and subcortical nuclei, to redundant networks and functioning, learning adapting systems, and eventually to cognition and more. Each level with its rules constrains the structure and function of the next, more complex ones. There are many examples of such rules. The Hodgkin-Huxley equations are the best known and among the oldest (Hodgkin and Huxley 1952). In principle, then, all the levels of brain organization should eventually become expressible in terms of mathematical relationships, and that would constitute a brain theory, or model.

Computers

A decade or two ago, the idea that an inestimable number of simultaneous non-linear equations could represent a theory of brain organization, if dreamt of by a few, seemed such an unlikely proposition that it merited no more than a passing *frisson*. There were two fundamental problems: how to make the calculations, and how to amass the data on which to make them. The first problem is largely solved, at least in principle and partly in practice. The most powerful super-computers currently available are at the peta-flop level (<http://www.top500.org/list/2013/06/>).

The IBM roadmap predicts the production of an exascale computer around 2018 (1×10^{18} flops/s). Extrapolating today's Blue Brain Project numbers, exascale is probably the minimum required to simulate the entire brain. This level of performance is just sufficient for the simultaneous computation of the present estimate of the number of equations needed to provide a first holistic version of a brain model, one that instantiates the nonlinear interactions that give rise to the emergent properties of living brains. As to data storage, this is a practical problem that has effectively been solved by cloud computing and distributed storage with appropriate addressing; it is data analysis and aggregation with efficient database queries that are challenges at this scale.

Data and Data Mining

Clinical scientists are used to dealing with highly controlled, “clean” data sets, despite the messy nature of their observational constructs. Hence their data sets are often small, precious and closely guarded, being a critical part of the discovery process. This mind set is invalidated by advances in data mining algorithms that have become commonplace in industry (banking, nuclear power, air transportation, space and meteorology, to name but a few) (http://en.wikipedia.org/wiki/Data_mining).

Such algorithms identify patterns in big data that are characterized by invariable clusters of (mathematical) rules. In other words, they are rule-based classifiers. They offer a potential escape from the world of correlations into the world of causes. However, strictly, rule-based classification generates correlations, not causality (although it depends on how narrowly causality is defined). It shows what occurs together but not what causes what. Homogeneous clusters are useful for disease signatures, but for treatments causality will have to be understood by integration of knowledge and simulation results from genetics, biochemistry, physiology and medical description into randomized experiments (Fig. 1).

These powerful and computer-sensitive, data-hungry algorithms often use novel mathematics. They have been developed because the new generations of computers can verify and validate them. They deal with multivariate and “dirty” data, missing data, textual or semantic data and data from different sources or with different ranges. They can work in non-linear, non-Euclidean, non-stochastic, high-dimensional spaces (Loucoubar et al. 2013). Others are more statistically based, such as machine learning techniques. Some attempt to exhaustively test all possible models describing the data to discover the most parsimonious set that explains them. Which will be the best tools and methods for use in the clinical neurosciences is not yet clear, but one can be sure that data mining will generate many hypotheses for testing! And so the perspective emerges that the comprehension of brain organization and the causes of brain disease are not to be found by a reductionist approach alone but by a combination of hypothesis falsification that follows a constructivist, simulation-based approach using novel classifiers working on large amounts of real biological data.

Simulating the Brain

An initial proof-of-concept program has recently communicated very encouraging results. The Blue Brain Project (<http://bluebrain.epfl.ch>) at the Brain Mind Institute of the Ecole Polytechnique Fédérale de Lausanne (EPFL) took as its starting point data on the functionality of ion channels and their distributions along axons and dendrites of different neural types (Peck and Markram 2008; Khazen et al. 2012). Proceeding with a simulation-based approach, using biological data about matters

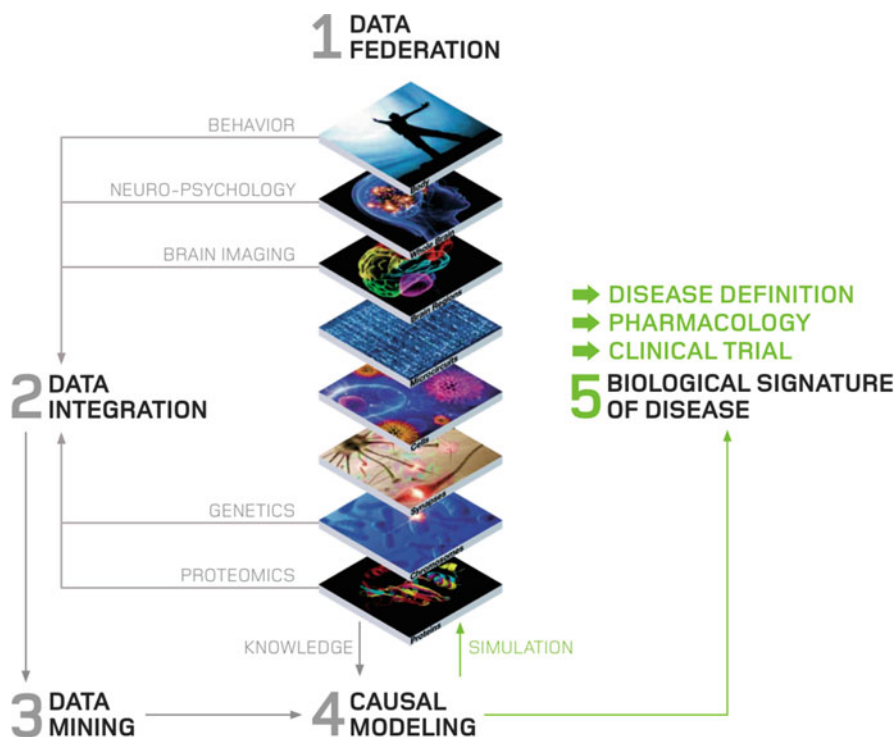


Fig. 1 Schematic representation of steps in applying modern informatics to clinical neuroscience. The Human Brain Project (HBP) aims to collect, explain and simulate the functions of the human brain at different levels of hierarchical complexity. Within the HBP framework, a strategically feasible approach to understand brain diseases is depicted in this figure. The idea is to federate (1) and integrate (2) the data, thus making use of an abundance of biological information from the different levels of brain organisation. Data mining (3) will be used to extract sets of rules that constitute definitions of homogeneous groupings of patients or subjects. Causal modelling with new data (real or simulated) will be performed for external validation (4), which will complete the process of defining (5) the biological signatures of diseases. The signatures of diseases will constitute the basis for a new, biologically determined nosology that should facilitate drug target identification and selection of homogeneous groups of patients for clinical trials as well as simulation of the effects of pharmacological treatment and secondary event profiles

such as cortical volume, the distribution of cortical layers, the distribution of capillaries, the variation in numbers and distributions of morphological and functional types of neurons in the various cortical layers, and statistical data on the probability of connections between different cell types, it built a preliminary model of a rodent cortical column using an IBM Blue Gene/Q computer. A correspondence of functionality and morphology between the model and *ex vivo* slices of brain tissue has been demonstrated (Ramaswamy et al. 2012). Predictions about the distribution of synaptic connections (Hill et al. 2012) and the occurrence of spontaneous activity (in the gamma band) have also been made that in themselves constitute strong hypotheses for further empirical verification.

Data Provenance

The computing power needed for the extension of such a project to whole brains is now within our reach. Data provenance remains a problem. In the research domain, there are 30 years of data described in millions of scientific papers lodged in repositories such as the National Library of Science in Washington DC. There are many basic science laboratory databases, often publically funded, held in universities and research laboratories around the world. These data have often been used once and exist for archival reasons alone. In the clinical field, there are databases in each hospital that contain clinical and diagnostic information on innumerable patients. Again, the data are used for the benefit of an individual and are normally kept for medico-legal reasons or as a baseline for returning patients. In countries with socialized medicine, these data are paid for by taxes and so, at least in part, belong to the public. This mass of legacy data represents an enormous, untapped research resource. How can such heterogeneous data be usefully exploited?

Real-time data addressing is a fact of life for anyone who uses the Internet and a search engine today. Therefore, in principle, the infrastructure and software are available. It remains to be seen whether specialized integrated hardware and software infrastructures will become acceptable to hospitals and researchers for scientific activity. Issues such as privacy protection in the context of anonymization are technically solvable and already acceptable on the grounds of proportionality (the potential benefit to members of society as a whole, compared to the potential risk to an individual) in worlds such as those of Internet banking and crime prevention (<http://www.scienceeurope.org/uploads/Public%20documents%20and%20speeches/ScienceEuropeMedicalPaper.pdf>; but see Gymrek et al. 2012). Following the CERN model, asking for scientists' data in return for giving them access to many other databases should be a huge incentive, especially since it will accelerate the process of scientific discovery by increasing the efficiency of data usage. The acceptability of such systems will depend on their ability to avoid displacement and corruption of source data, which is already a practical possibility (Alagiannis et al. 2012; Fig. 2). The advantage to society is that taxpayers will contribute to medical research at no extra cost while benefiting from its fruits. In other words, every datum collected in the course of standard medical care will also serve to promote medical and health research based on big data (Marx 2013).

Disease Signatures

Initially, the strategy is to federate data through a dedicated, connected infrastructure and then to integrate them appropriately so that they can be mined for answers to specific questions. In all cases the results will relate to groups and not to individuals, so guaranteeing an appropriate and proportionate degree of privacy.

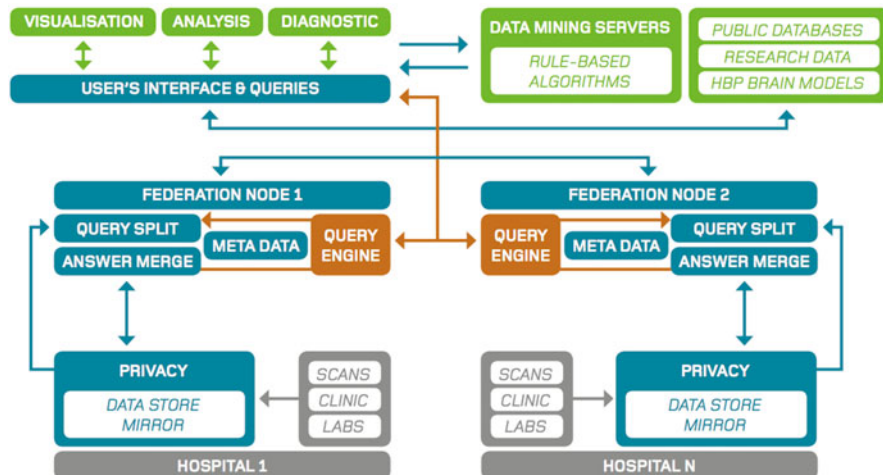


Fig. 2 Schematic describing the clinical neurosciences big data infra-structure. In the context of the Human Brain Project, research will be undertaken based on distributed processing of medical informatics infrastructures. The Medical Informatics Platform will provide a software framework that allows researchers to query clinical data stored on hospital and laboratory servers, without moving the data from the servers where they reside and without disproportionately compromising patient privacy (in situ querying). Tools and data queries will be made available to a participating community via a web-based technology platform adapted for neuroscientific, clinical, genetic, epidemiological and pharmacological users. The information made available will include brain scans of various types, data from electrophysiology, electroencephalography and genotyping, metabolic, biochemical and hematological profiles and also data from validated clinical instruments. Tools will be designed to aggregate data for analysis by state of the art high-performance computing that automatically provides a basic descriptive statistical overview as well as advanced machine learning and discovery tools

“The Human Brain Project,” awarded one billion euros in a European Commission Flagship of Enterprise and Technology competition in 2013, seeks to use this strategy in its medical informatics division (<http://www.humanbrainproject.eu/#>). One type of question will involve identifying groups of patients who show identical patterns of biological abnormality based on the results of clinical investigation. These patterns, called “disease signatures,” will comprise sets of causative features including clinical findings, results of validated questionnaires of mood and emotion, brain images of various types, electrophysiological recordings, blood tests, genotypic characteristics, and protein composition of cerebrospinal fluid or blood. To obtain maximal differentiation and sensitive discrimination between different diseases, the strategy will be to use data from as wide and inclusive a range of brain diseases (both neurological and psychiatric) as possible. This approach runs directly counter to standard techniques of epidemiology based on tightly defined syndromes or single characteristics, such as a unique pattern of single nucleotide polymorphisms or protein expression, by seeking to understand and resolve the one syndrome—multiple mutations and one mutation—multiple phenotypes problems. The disease space, sampled in multiple dimensions, each of which is described by a

specific vector of biological variables, will provide a new diagnostic nosology that is in principle quantitative and expressed by a complete, exclusive set of characteristic clinical features and results.

In the context of a medical consultation, a doctor might take a set of measurements and order a set of tests to provide a results vector, which can be presented to a database for matching to disease type, a clear step towards personalized medicine. Biologically characterized diagnostic criteria will facilitate drug trials in that diagnostic ambiguities in control and disease cohorts will be drastically attenuated, leading to small groups with reduced error variances and adequate power for drug discovery in humans. In dementia, as mentioned earlier, the error in AD diagnosis approaches 30 %. Certain aged normal people have a degree of AD-related pathological change, which is compensated for at the behavioral or cognitive level. It is claimed that 39 % of elderly subjects supposed to be normal show AD pathology post mortem (Schöls et al. 2012). Twenty percent of 80-year-old adults have some form of recognizable cognitive decline, so the error variance in currently constituted normal control groups may also be substantial. Clinical trials with groups that are as inhomogeneous as these are likely to fail, even with specifically targeted drugs. A search for preclinical abnormality in populations may lead to a definition of types of “normality” in large enough data sets, and the dementias may become more usefully defined by shared clinical and biological characteristics.

Data Mining and Medical Data

One data mining tool—Hypercube©—has already been used in medical research (Loucoubar et al. 2011). We have preliminary data with this algorithm on a set of 200 patients from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (<http://www.three-city-study.com/les-publications-scientifiques-de-l-etude-des-trois-cites.php>) and also from a subset of 500 elderly subjects from the 3 Cities study (<http://adni.loni.ucla.edu>) and associated image-genetics-clinical-psychology cohorts followed in France for 10 years. Our analyses are somewhat flawed, in that the entire disease space is not sampled and the numbers of patients are pitifully small (though we now have over 6000 donated datasets from the same sources and from the pharmaceutical company Sanofi-Aventis), but encouraging patterns have emerged. Of the subjects in the first dataset, 199 of 200 fell into six distinct subgroups on the basis of “disease signature.” In the second, where substantial genotyping data were also available, separate, normal-aged groups can be distinguished from a number of groups associated with cognitive decline. The largest of the latter includes APP and ApoE4 in its “disease signature.” Of great interest will be secondary phenotyping, returning to groups of patients with the same “disease signature” to identify specific clinical characteristics or variability in them with factors such as age, which will give further insight into how brain diseases manifest (Fig. 3).

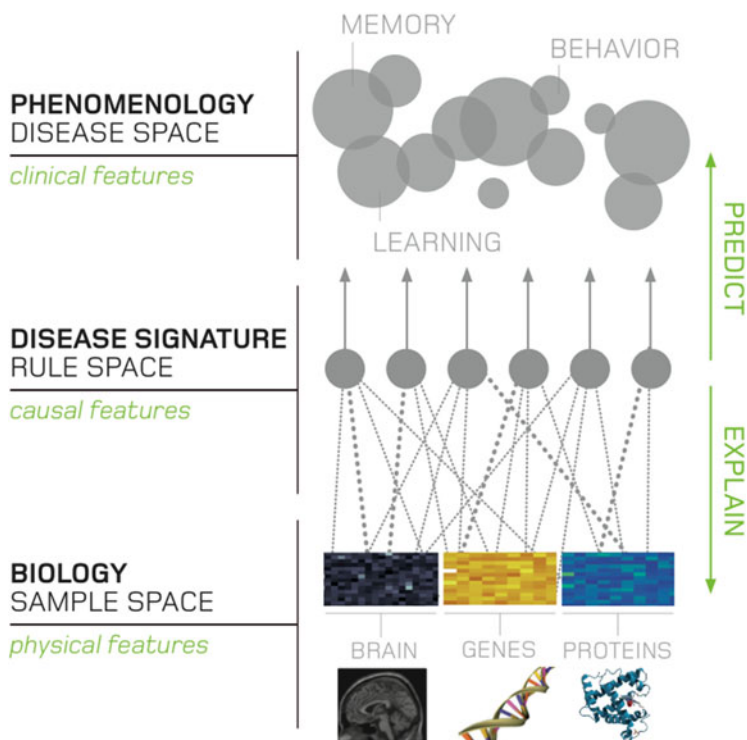


Fig. 3 Theoretical schema describing the relationship between different levels of description and the role of the disease signature in relating biology to phenomenology. The biological signatures of diseases are deterministic mathematical constructs that aim to describe both variability at the phenomenological level (clinical features with symptoms and syndromes) and at the biological level (genetic, proteomic, etc.). The key property of a biological signature of disease is that it accounts for the fact that a symptom of brain dysfunction can be due to many biological causes (one-to-many symptom mapping) and that a biological cause can present with many symptoms (many-to-one symptom mapping). In reality, the situation is often one of many-to-many mappings between symptoms and biological causes. With advanced computing power, nearly exhaustive searches of a data space can be performed to identify sets of rules that describe homogeneous populations, to explain their biological data and to predict the pattern of symptoms

Human Brain Project

The Human Brain Project has, in addition to a medical informatics division, a basic neuroscience component that is charged with creating an in silico blueprint (model) of the normal human brain. Replacement of normal biological characteristics in such a model by disease-associated values should, if correct, give an idea after propagation through the model of what associated functional or structural changes to expect. Likewise, modifications of parameters induced by a neuromodulator or other factor should provide ideas about the spectrum of both desired and undesired effects of any such medication (Harpaz et al. 2013). It may be worth enlarging this

perspective to system-based approaches, too (Zhang et al. 2013). In a real sense the normal brain simulation program and the medical informatics effort will serve to test each other in a cycle of repeated virtuous iteration until adequate accuracy can be achieved for medical practice.

Europe has provided funds for a major coordinated effort in this field, supported by leading edge computer science and technology, which has its own agenda of using knowledge about human brain organization to inspire novel chip and computer architectures. The aim is to move on from von Neumann digital binary machines to neuromorphic probabilistic architectures that are much more energy-efficient (Pfeil et al. 2013; Indiveri et al. 2011). The vision described here is broad but practical. Its implementation will demand new competencies in medical researchers and doctors, greater cross-disciplinary collaboration (along the lines pioneered by physicists in CERN) and major changes in culture and practice.

Clinical Neuroscience-Related Big Data Initiatives

The scientific world is taking on the challenge faced by clinical neuroscience to create a culture and foster competences that will be needed for the effective use of big data research (see Box 1 for more details). Examples include BioMedBridges, a joint effort by ten European biomedical sciences research infrastructures in which the project partners will develop a shared e-infrastructure—the technical bridges—to allow interoperability between data and services in the biological, medical, translational and clinical domains; One Mind, which has the vision of a technology-enabled data-sharing community focused on psychiatric disease and brain injury, brought together through a federated data platform; the Allen Brain Atlas, a growing collection of online public resources integrating extensive gene expression and neuroanatomical data, including that of humans, complete with a novel suite of search and viewing tools; ELIXIR, which unites Europe’s leading life science organizations in managing and safeguarding the massive amounts of data being generated every day by publicly funded research; and ENIGMA (Enhancing Neuro-Imaging Genetics through Meta-Analysis), which brings together researchers in imaging genomics to understand brain structure and function based on MRI, DTI, fMRI and GWAS data.

Box 1. Big data initiatives: a selection

Allen Brain Atlas—<http://www.brain-map.org>

BioMed Bridges—<http://www.biomedbridges.eu>

One Mind—<http://1mind4research.org/about-one-mind>

Elixir—<http://www.elixir-europe.org>

Enigma—<http://enigma.loni.ucla.edu/about/>

European Bioinformatics Institute—<http://www.ebi.ac.uk>

International Neuroinformatics Coordinating Facility—<http://www.incf.org/about>

Machine Learning—http://en.wikipedia.org/wiki/Machine_learning

Data mining—http://en.wikipedia.org/wiki/Data_mining



Changing the Culture

Far-seeing higher educational establishments such as the EPFL have been developing strategies of recruitment and faculty development that bring engineering and ICT together with life and clinical sciences in preparation for such a revolution.

The public will need to be convinced of the privacy issues, and researchers will need to acknowledge that it is ideas and not just data that generate Nobel Prize-winning work. Finally, politicians and industrialists will need to be convinced that there are substantial efficiency savings to be made by preventing the endless repetition of underpowered studies with unreproducible results that characterize much of present-day life science. They will presumably be open to exploiting the added value that federating data offers at no extra cost and to the business opportunities that arise from developing, installing and maintaining local infrastructures to feed big data-based medical and health sciences research on a global scale (Hood and Friend 2011).

Acknowledgments This work benefited from funding by the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 604102 (Human Brain Project).

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Alagiannis I, Borovica R, Branco M, Idreos S, Ailamaki A (2012) NoDB: efficient query execution on raw data files. In: ACM SIGMOD international conference on management of data, ACM, 978-1-4503-1247-9/12/05
- Beach TG, Monsell SE, Phillips LE, Kukull W (2012) Accuracy of the clinical diagnosis of Alzheimer Disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *J Neuropathol Exp Neurol* 71:266–273
- Brahham DC (2008) Crowdsourcing as a model for problem solving: an introduction and cases. *Convergence Int J Res New Media Technol* 14:75–90
- Brownstein JS, Freifeld CC, Reis BY, Mandl KD (2008) Surveillance sans frontieres: internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med* 5:e151. doi:10.1371/journal.pmed.0050151
- Chiang G-T, Clapham P, Qi G, Sale K, Coates G (2011) Implementing a genomic data management system using iRODS. *BMC Bioinformatics* 12:361
- CMS Collaboration, Chatrchyan S, Khachatryan V, Sirunyan AM, Tumasyan A, Adam W, Aguilo E, Bergauer T, Dragicevic M, Erö J, Fabjan C, Friedl M, Frühwirth R, Ghete VM, Hammer J, Hoch M, Hörmann N, Hrubec J, Jeitler M, Kiesenhofer W, Knünz V, Kramme M, Krättschmer I, Liko D, Majerotto W, Mikulec I, Pernicka M, Rabbaran B, Rohringer C, Rohringer H, Schöfbeck R, Strauss J (2012) Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys Lett B* 716:30–61

- Cross-Disorder Group of the Psychiatric Genomics Consortium (2004) Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381:1371–1379
- Goldstein DB (2009) Common genetic variation and human traits. *N Engl J Med* 360:1696–1698
- Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y (2012) Identifying personal genomes by surname inference. *Science* 339:321–324
- Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C (2013) Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin Pharmacol Ther*. doi:[10.1038/clpt.2013.125](https://doi.org/10.1038/clpt.2013.125)
- Hay A, Tansley S, Tolle K (2009) *The fourth paradigm: data-intensive scientific discovery*. Microsoft, Redmond, WA. ISBN 978-0-9825442-0-4
- Hill SL, Wang Y, Riachi I, Schurmann F, Markram H (2012) Statistical connectivity provides a sufficient foundation for specific functional connectivity in neocortical neural microcircuits. *Proc Natl Acad Sci USA* 109:E2885–E2894
- Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *Physiology* 117:500–544
- Hood L, Friend SH (2011) Relevance of network hierarchy in cancer drug-target selection. *Nat Rev Clin Oncol* 8:184–187
- Indiveri G, Linares-Barranco B, Hamilton TJ, van Schaik A, Etienne-Cummings R, Delbruck T, Liu S-C, Dudek P, Häflicher P, Renaud S, Schemmel J, Cauwenberghs G, Arthur J, Hynna K, Folorosele F, Saighi S, Serrano-Gotarredona T, Wijekoon J, Wang Y, Boahen K (2011) Neuromorphic silicon neuron circuits. *Front Neurosci* 5:73. doi:[10.3389/fnins.2011.00073](https://doi.org/10.3389/fnins.2011.00073)
- Kertesz A, McMonagle P, Blair M, Davidson W, Munoz DG (2005) The evolution and pathology of frontotemporal dementia. *Brain* 128:1996–2005
- Khazen G, Hill SL, Schuermann F, Markram H (2012) Combinatorial expression rules of ion channel genes in juvenile rat (*Rattus norvegicus*) neocortical neurons. *PLoS One* 7:e34786. doi:[10.1371/journal.pone.0034786](https://doi.org/10.1371/journal.pone.0034786)
- Loucoubar C, Paul R, Huret A, Tall A, Sokhna C, Trape J-F, Ly AB, Faye J, Badiane A, Diakhaby G, Sarr FD, Diop A, Sakuntabhai A, Bureau J-F (2011) An exhaustive, non-Euclidean, non-parametric data mining tool for unraveling the complexity of biological systems—novel insights into malaria. *PLoS One* 6:e24085. doi:[10.1371/journal.pone.0024085](https://doi.org/10.1371/journal.pone.0024085)
- Loucoubar C, Grange L, Paul R, Huret A, Tall A, Telle O, Roussilhon C, Faye J, Diene-Sarr F, Trape JF, Mercereau-Puijalon O, Sakuntabhai A, Bureau JF (2013) High number of previous *Plasmodium falciparum* clinical episodes increases risk of future episodes in a sub-group of individuals. *PLoS One* 8:e55666. doi:[10.1371/journal.pone.0055666](https://doi.org/10.1371/journal.pone.0055666)
- Marks P (2007) Massive science experiments pose data storage problems. *New Sci* 196:28–29
- Marx V (2013) The big challenges of big data. *Nature* 498:255–260
- Peck C, Markram H (2008) Identifying, tabulating, and analyzing contacts between branched neuron morphologies. *IBM J Res Dev* 52:43–55
- Pfeil T, Grubl A, Jeltsch S, Muller E, Muller P, Petrovici MA, Schmuker M, Bruderle D, Schemmel J, Meier K (2013) Six networks on a universal neuromorphic computing substrate. *Front Neurosci* 7:11. doi:[10.3389/fnins.2013.00011](https://doi.org/10.3389/fnins.2013.00011)
- Rajasekar A, Moore R, Hou CY, Lee CA, de Torcy A, Wan M, Schroeder W, Chen SY, Gilbert L, Tooby P, Zhu B (2010) iRODS primer: integrated rule-oriented data system. *Synthesis lectures on information concepts, retrieval, and services*. Morgan & Claypool, San Rafael, 143p
- Ramaswamy S, Hill SL, King JG, Schurmann F, Wang Y, Markram H (2012) Intrinsic morphological diversity of thick-tufted layer 5 pyramidal neurons ensures robust and invariant properties of in silico synaptic connections. *J Physiol (Lond)* 590:737–752. doi:[10.1113/jphysiol.2011.219576](https://doi.org/10.1113/jphysiol.2011.219576)
- Schöls L, Bauer P, Schmidt T, Schulte T, Riess O (2012) Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. *Lancet Neurol* 3:291–304
- Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezchnikov AA, Zhang C, Xie T, Tran L, Dobrin R, Fluder E, Clurman B, Melquist S, Narayanan M, Suver C, Shah H, Mahajan M, Gillis T, Mysore J, MacDonald ME, Lamb JR, Bennett DAB, Molony C, Stone DJ, Gudnason V, Myers AJ, Schadt EA, Neumann H, Zhu J, Emilsson V (2013) Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* 153:707–720

Index

A

Auditory novelty signal, 90–93

B

Blue brain project, 160–162
Bottom-up, 44, 46, 54–55, 90, 136
Boundary vector cell (BVCs), 3
Brain and behavior, 30–31

C

Clinical neuroscience, 157–169
Cognitive map, 2
Collective predictive belief, 54
Combinatorics in grid cells, 64–67
Computer science, 157–169
Computing through multiscale complexity, 43–55
Consciousness, 92–95, 108–109, 111, 112, 115–117, 139–141, 144–145, 148
Cortical evolution, 23–31

D

Data mining, 161, 162, 165–166, 168
Data provenance, 163
Decision making, 81–83
Discretization of the entorhinal grid map, 62–64
Disease signatures, 161, 163–166
Dopamine, 82, 83
Dynamics of conscious perception, 85–95

E

EEG. *See* Electro-encephalography (EEG)
Electro-encephalography (EEG), 108, 133–135, 164
Embryonic development, 24
Emergence, 46, 52, 55
Entorhinal cortex, 5, 28, 59–76

F

Federating what we know about the brain, 157–169
Flow of information, 35–40

G

Genome-wide association study (GWAS), 158, 167
Gestalt, 46, 51–54
Gist extraction, 100–102
Global neuronal workspace (GNW), 93, 95
Grid cell, 3, 4, 7, 59–76
Grid spacing, 4, 62–63, 65, 69–71

H

Head-fixed tactile decision, 36–37, 39
Hippocampus, 2, 4–8, 11–14, 24–25, 30–31, 59–76, 102, 123, 127
Homeostasis, 24, 49, 54–55, 99–100, 104–105
Horizontal connectivity, 27–29, 49, 51, 53–54
Human brain project (HBP), 162, 164–167

I

Immersion, 45–51, 54
 Integration of new with old memories, 101

L

Local field potential (LFP), 5, 7, 8, 12–13, 63, 81, 118, 123–125, 127, 130–135

M

Magneto-encephalography (MEG), 85–90, 92, 108–109, 135
 Memory, 2, 6, 13–15, 30–31, 35, 40, 64, 68, 74, 100–105, 114
 consolidation, 100–102, 104
 Metastability, 93–95
 Motor cortex, 35–36, 38
 Motor planning, 35–40
 Multivariate decoding method, 88–90, 95

N

Neural word, 10
 Neuronal doctrine, 44

O

Oblique effect in grids, 74–76
 Optogenetic, 36, 82, 83, 114
 Oscillation, 5, 7, 28, 63, 127, 136

P

Parahippocampal system, 67–69
 Physiological adaptation of the reptilian brain, 29–30
 Place cell, 3, 4, 7–10, 12, 14, 59–62, 64–68, 74
 Place field, 2–4, 6–14, 64–67, 131
 Plasticity, 4, 9, 102, 104–105, 137
 Prefrontal cortex, 81, 82, 90, 136
 Premotor dynamics, 38–40

R

Receptive fields (RFs), 43–54, 117
 REM sleep, 13, 99–102, 104
 Reptilian cortex, 23–31

S

Segmentation of space by goals and boundaries, 1–15
 Shearing-induced asymmetries in grid cells, 71–74
 Simulating the brain, 161–162
 Sleep, 6, 13, 92, 99–105, 108–109, 112, 115, 136
 in sleep, 99–100, 102–105
 Spatial
 code, 5–8, 10–11, 14–15
 map, 5, 59–76
 navigation, 2
 representation, 4, 67–68, 74
 Striatum, 12, 81–83
 Striosome, 82, 83
 Synapse, 26, 100–105, 118, 122–126, 128, 129, 136–137, 139, 140
 Synaptic
 consolidation, 102
 down-selection, 99–105
 Synaptic homeostasis hypothesis (SHY), 104–105
 Synaptic imprint of mesoscopic immersion, 46–51
 Syndromic diagnosis, 159
 System-level consolidation, 102

T

Tactile decision in mice, 35–40
 Temporal decoding, 87, 88, 90–92, 94, 95
 Temporal generalization method, 85–95
 Theory of the brain, 159–160
 Theta rhythm, 2, 5, 63
 Top-down, 45–46, 52–55, 90, 92
 Transcranial magnetic stimulation, 103–104

V

Vertical connectivity, 25–27
 Visual brain, 43–55